ELECTROENCEPHALOGRAPHIC MEASURES OF AUDITORY PERCEPTION IN DYNAMIC ACOUSTIC ENVIRONMENTS

AMANDA R. MCMULLAN
Bachelor of Science, University of Lethbridge, 2011

A Thesis
Submitted to the School of Graduate Studies
of the University of Lethbridge
in Partial Fulfillment of the
Requirement for the Degree

MASTER OF SCIENCE

Department of Neuroscience
University of Lethbridge
LETHBRIDGE, ALBERTA, CANADA

Dedication

To my grandfather, David Masters, for his unwavering interest in my happiness, my education, my life, and my future.

Abstract

We are capable of effortlessly parsing a complex scene presented to us. In order to do this, we must segregate objects from each other and from the background. While this process has been extensively studied in vision science, it remains relatively less understood in auditory science. This thesis sought to characterize the neuroelectric correlates of auditory scene analysis using electroencephalography. Chapter 2 determined components evoked by first-order energy boundaries and second-order pitch boundaries. Chapter 3 determined components evoked by first-order and second-order discontinuous motion boundaries. Both of these chapters focused on analysis of event-related potential (ERP) waveforms and time-frequency analysis. In addition, these chapters investigated the contralateral nature of a negative ERP component. These results extend the current knowledge of auditory scene analysis by providing a starting point for discussing and characterizing first-order and second-order boundaries in an auditory scene.

Acknowledgments

First and foremost I would like to thank my supervisor, Dr. Matthew Tata for his ability to see my potential, even when I could not. I was fortunate to have a supervisor who was patient enough to help guide my passion, curiosity, and allow me to expand my understanding of not just neuroscience, but science and academia as a whole. I have gained valuable life experience from working with him over the past years and, thanks to him, will continue to foster my scientific mind.

I would also like to thank my committee members, Dr. Claudia Gonzalez and Dr. Arlan Schultz, and my thesis chair, Dr. Andrew Iwaniuk, for their input and guidance through the process of this graduate thesis.

I was fortunate to share a lab and research with many research assistants and fellow graduate students. Thank you to Dillon Hambrook, Karla Ponjavic-Conte, Scott Oberg, Andrew Butcher, Tom Rutherford, and Sheena MacInnis for listening to me, encouraging me, and helping me.

To my family, who has dealt with my somewhat absence through the past two years, I thank you. Especially, my mom, Myrna, for ensuring I had everything I needed to make it through; my dad and step-mom, Ian and Brenda, for their words of encouragement; and my aunt, Laura, for her devotion to my life and always being in my corner. To my friends who have dealt with my busy schedule and worked around it, thank you. Especially, Adam Dipinto, for his support, words of kindness, and reassurance; and Noël Matchett, for always ensuring I have someone to go to and a delicious meal to eat. I am and will always be grateful.

Table of Contents

List of Figures

List of Abbreviations

ANOVA – Analysis of Variance

EEG – Electroencephalography

EP – Evoked Power

ERP – Event-Related Potential

FDR – False-Discovery Rate

fMRI – Functional Magnetic Resonance Imaging

GBR – Gamma-Band Response

IP – Induced Power

IRN – Iterated Rippled Noise

ISI – Inter-Stimulus Interval

ITPC – Inter-Trial Phase Coherence

LFP – Local Field Potential

LORN – Lateralized Object-Related Negativity

MEG – Magnetoencephalography

MMN – Mismatched Negativity

ORN – Object-Related Negativity

PET – Positron Emission Tomography

POR – Pitch Onset Response

RMS – Root Mean Square

TSE – Time-Spectral Evolution

VBAP – Vector-Based Amplitude Panning

Chapter 1: Introduction

This thesis is about auditory scene analysis. The goal of the experiments described here was to explore and characterize neuroelectric correlates of the processes by which the brain perceives groups of acoustic signals as distinct auditory objects. Specifically, I tested the theory that different kinds of boundaries in the auditory world engage different neural mechanisms.

1.1 Visual and Auditory Scene Analysis

Arguably, one of the most complex tasks that the brain undertakes is that of scene analysis. This consists of parsing a scene for objects and interpreting the boundaries or edges between these objects. In both the visual and auditory domain, our perceptual systems are constantly dealing with stimuli that need to be separated from other objects and background, so that we can interact with them effectively. Objects in vision are intuitive and can be easy to describe. A visual object can be a ball, a truck, a tree, etc. Visual objects usually, but not always, have common and intuitive characteristics: they can be grasped or touched, they occlude the background or other objects behind them, and they move according to laws of physics. Importantly, they usually have surfaces that reflect or scatter light in ways that allow us to perceptually separate one object from another. These visual features include, for example, color, size, distance, and orientation.

By contrast, auditory objects are less intuitive, but no less perceptually real (Griffiths & Warren, 2004; Kumar *et al.*, 2007). As in vision, an auditory scene can contain many different acoustic events. However, unlike visual objects, auditory objects

are usually abstracted from physical objects. When a sound is heard, we can identify it and describe it in terms of its loudness, pitch, and timbre. We can say whether it is moving or stationary. We perceive it as rising or falling in pitch. Yet we do not usually imagine that we might reach out and touch the sound itself. An auditory object corresponds only abstractly with a physical object and it is subject to the physical laws of acoustics, not the Newtonian mechanisms that govern most behavior of visual objects. The auditory scene is dynamic in ways the visual scene simply is not. For example, most surfaces reflect sound such that acoustic events usually reverberate in auditory space (imagine a world in which every surface was a mirror). Most importantly, the auditory and visual systems differ in a fundamental way: whereas the visual system features a spatiotopic organization because of the image-forming lens and cornea of the eye, the auditory system is inherently tonotopic. The sensory surface of the auditory system, the basilar membrane, decomposes complex sounds into component frequencies and represents these spatially across its surface. The consequence is that the dynamics of objects in an auditory scene are encoded in the spectral evolution of sound.

Despite the dissimilarity between visual and auditory objects, we find a number of commonalities in the process by which the brain encodes and interacts with objects. For example, Bregman (1990) characterized the conditions under which 'acoustic events' are perceptually bound into what he called an 'auditory stream'. The resulting rules of auditory scene analysis bear strong resemblance to the gestalt grouping rules of visual scene analysis. For example, sound from a physical source does not usually change suddenly in several different dimensions at once. Thus, abrupt spectral and spatial changes indicate that an onset or offset of an auditory object has occurred. These abrupt

boundaries in time between 'regions' of different pitch, timbre, or spatial cues are the acoustic analogs of edges between objects in the visual world.

Vision science has extensively investigated the boundaries between visual objects and the background. A variety of visual cues are known to be encoded in the steps required to parse a visual scene. These include spatial discontinuities in brightness or luminosity that are encoded in the responses of visual neurons at the lowest levels of the visual system (Hubel & Wiesel, 1962). These luminosity boundaries have been termed *first-order* and they trigger robust electrical responses when they change. Later research turned to neuronal responses to other types of boundaries, such as isoluminant color (Shapley, 1990; Shapley, 2002), texture (Mareschal & Baker, 1998), and motion (Albright, 1992; Anderson, 1997). Boundaries in the visual scene that do not contain luminance or brightness discontinuities were termed *second*-order or *higher*-order.

Like visual boundaries, auditory boundaries usually involve a first-order feature – an increment or decrement in acoustic energy across a broad range of frequencies. However, auditory boundaries that do not involve a broad change in energy have recently come under scrutiny. *Second*-order boundaries in the acoustic scene include pitch discontinuities (Ungan *et al.*, 1989; Johnson *et al.*, 2003; Krumbholz *et al.*, 2003; Seither-Preisler *et al.*, 2004; Seither-Preisler *et al.*, 2006a), discrepancies in harmonic series (Alain *et al.*, 2001b; Alain *et al.*, 2002), changes in motion trajectory (Butcher *et al.*, 2011), and changes in interaural correlation (when both ears are receiving the same sound signal, but one has a time-lag) (Chait, 2005; Chait *et al.*, 2007a). In these situations, investigators have focused on the brain activity that underlies the encoding of the boundary itself, usually with an effort to uncouple such signals from the more general

and well-studied responses of the brain to first-order changes in sound energy. To understand the progress made to date, and to introduce the technique used in the research described in this thesis, we turn now to an overview of electroencephalography (EEG) and its application to the field of auditory cognitive neuroscience.

1.2 Electroencephalography Methodology

Neurons exchange information through a series of electrical and chemical signals. The biophysical correlate of this process is a time-varying change of the electrical charges that surround the neural cell membrane. When these electrochemical signals arrive at the cell bodies of neurons, the resulting movement of charges across the membrane results in a transient *graded potential*, which is a weak electrical field around the cell body. This graded potential is distinct from the well-known action potential, which ultimately conveys signals to other neurons, but is too weak and transient to be measured at the scalp. The graded potential around any neural cell body is too weak to be measured at any appreciable distance from the cell. However, the potentials from large groups of neurons acting coherently sum to form a *local field potential* (LFP), which is conducted through the intracranial volume to the scalp (Whittingstall & Logothetis, 2009). It is this aggregate LFP that is thought to form the basis of the familiar scalp-recorded EEG.

Many neuroimaging techniques have been used to study brain responses to visual and auditory stimuli. Each technique is associated with a list of capabilities and limitations that must be considered with respect to the kind of information that can be gleaned about brain activity. The technique selected depends greatly on the nature of the

phenomena under study and the nature of the questions being asked. EEG and its magnetic counterpart, magnetoencephalography (MEG), have been used extensively to investigate temporal aspects of brain activity on the timescale of milliseconds. These techniques are therefore the preferred strategies to investigate rapidly changing brain activity. However, although EEG has high temporal resolution, it has poor spatial resolution. An array of EEG sensors typically includes as many as 128 electrodes (dense-array EEG), which allows for the creation of a topographic 'map' of voltage across the scalp. While, the voltage distribution can be used to estimate neural sources, it is limited by distortions that occur as the electrical signal spreads through the intracranial volume. When spatial precision is required, cortical activity of the brain is resolved best by metabolic techniques such as functional Magnetic Resonance Imaging (fMRI) and Positron Emission Tomography (PET). EEG was chosen for the studies described in this thesis because we placed emphasis on resolving the fast-changing dynamics of the brain, rather than on mapping the precise structures engaged in auditory processing.

EEG and MEG are particularly useful in investigating cognitive processes that may vary when a stimulus is manipulated or modified in systematic ways. During an experiment, EEG signal is continuously recorded. Of interest to many cognitive neuroscientists is the time just before or after a stimulus appears in a scene. Because of this temporal reference point, the subsequent electrical potentials are referred to as *event-related potentials* or ERPs. These potentials exhibit a close relationship in time to a stimulus or other sensory or motor events. Analysis of ERP waveforms involves extracting the EEG data recorded before and after an event of interest and then averaging across repetitions of trials. This is typically done for several individuals and then 'grand

averaged' across participants to produce ERP waveforms such as the ones that appear in the following chapters. Thus, the grand averaged ERP reveals similarities across participants, although it necessarily does not reflect individual differences in ERPs.

One common approach to the investigation of brain activity using the ERP technique is to measure and interpret variations in the timing and amplitude of peaks, both positive and negative, in the resulting waveforms. These peaks, sometimes also called *components*, are often labeled based on their polarity and latency within the waveform (Luck, 2005). For example, the N1 or N100 component is a negative deflection occurring approximately 100ms after stimulus onset. It is important that these labels are not necessarily linked with an underlying signal or a unique function. For example, both auditory and visual stimuli elicit a negative N1 and a positive P2, but these components do not exhibit a relationship to one another and do not necessarily reflect the same cognitive operations (see Luck, 2005 for review on EEG terminology). In the studies described below, particular emphasis is placed on three such peaks in the ERP waveform: a positive deflection at around 90ms, known as the P90; the negative N1 deflection at approximately 100-150ms; and the positive P2 at about 200ms. Associated with these are subtle variations or differences between peaks under certain conditions. For example, Chapter 3 describes a negative deflection slightly later than the typical N1 peak that varies in amplitude at certain electrodes depending on whether the stimulus appeared on the left or right side of the listener's midline.

The EEG is comprised of an ongoing oscillatory signal from which the ERP waveform can be computed. The amplitude (or power) of this oscillatory signal is not always distributed uniformly across the range of component frequencies. Event-related

perturbations in the oscillatory signal results in transient changes in power at certain frequencies. These perturbations of power are typically analyzed in reference to a pre-stimulus baseline and expressed as a percentage change. Signal that is jittered in time and/or phase is called induced. In order to investigate the induced power of EEG signal, complex demodulation is carried out. This allows the signal to be transformed into time-frequency space. Complex demodulation allows one to investigate components on a trial-by-trial basis. The EEG signal is demodulated using a discrete algorithm that separates and computes power at multiple frequencies over time (Hoechstetter *et al.*, 2004). Once complete, the single-trial data is averaged for each individual subject and then grand-averaged across those subjects. In a similar way, the phase of the response to a stimulus is also computed from the demodulation and is known as inter-trial phase coherence. It is the measure of how consistent the phase of the response is over consecutive trials. When the EEG signal exhibits substantial phase-locking to a stimulus event we refer to it as *evoked* power; when the EEG exhibits a change in power but the signal is not phase-consistent over successive trials we refer to it as *induced* power.

The time-frequency analysis of the EEG allows evaluation of signal at multiple frequency ranges. The frequency spectrum is divided into distinct frequency bands. These bands are called delta (0-4Hz), theta (4-8Hz), alpha (8-12Hz), beta (12-30Hz), and gamma (>30Hz). In EEG research, the alpha and theta frequency bands are often of interest because sensory evoked potentials occur at this frequency range. Oscillations at these frequencies are important for a number of cognitive tasks. Theta oscillations have been associated with memory encoding and processing and alpha oscillations have been linked to perception, attention, and memory (see Sauseng & Klimesch, 2008 for review).

As time-frequency analysis has become more prominent, cognitive processes have been associated to other frequency bands. For example, gamma oscillations were first investigated in regard to visual awareness and are now thought to be related to the binding of information across disparate regions of the cortex (Singer & Gray, 1995). More recently, induced gamma has been associated with information encoding and retrieval (Tallon-Baudry & Bertrand, 1999). At this frequency band, the distinction between evoked and induced power has been especially investigated. Gamma band oscillations are fast and have to occur with high phase locking in order to appear as evoked rather than induced signal. Furthermore, gamma band oscillations are small in amplitude, which makes it difficult to be analyzed with human EEG. Exogenous (bottom-up processing due to sensory stimuli) and endogenous (top-down processing due to cognitive processes) modulation of the gamma band response (GBR) have become of increasing interest and continue to be investigated (see Herrmann *et al.*, 2010 for review). Gamma band activity in the context of this thesis will be discussed further in Chapter 2.

1.3 Auditory Pathways

In order to better understand how the brain interprets incoming auditory stimuli, an understanding of the auditory pathway is needed. The auditory pathway takes the changes in sound pressure and turns them into neural signals. When a sound is produced, the sound waves travel to the eardrum, causing it to vibrate. These vibrations are in tune with the incoming stimuli. These vibrations in turn move a series of small bones in the middle ear. These bones carry the vibrations to fluid-filled cochlea in the inner ear. There are small hairs, called cilia, inside the cochlea that become stimulated along the basilar

membrane of the cochlea. The cilia are attached to auditory nerves so that when the cilia are stimulated, the nerves send signals to the brain. Cilia at the base of the cochlea are activated by high-frequency sounds and cilia at the far end of the basilar membrane are activated by low-frequency sounds. It is important to note that more complex sounds contain sound at multiple frequency ranges and these in turn stimulate more than one area on the basilar membrane.

Before the auditory information makes it to the cortex, there are a number of subcortical relays that the information passes through. From the cochlear nerves, output signal is sent to the superior olivary nuclei, the inferior colliculus, and the medial geniculate nuceli located in the thalamus. As with other sensory systems, the medial geniculate sends information to primary sensory cortex. In the case of an auditory stimulus, the information is sent to the primary auditory cortex, which is located in the superior part of the temporal lobe.

Auditory perception identifies the content of an acoustic stimulus and it also localizes sounds in space. In order to decipher an incoming sound, frequency data is crucial. Neurons throughout the auditory pathway have frequency tuning. While individual cells are unable to give precise frequency information, they are able to convey coarse coding. This means that auditory perception depends on the integration of multiple cells. It is likely that this integration comes from tonotopic maps, which show a correspondence of a neurons frequency tuning as well as its location. This helps for the perception of what the sound present is. However, localization of an acoustic signal requires integration of both ears. One of these integrations is called interaural time, which is the difference in when a sound reaches each of the two ears. When there is an

asymmetry in the arrival time at the two ears. Another integration of the two ears is called, interaural level or interaural intensity, which is the difference in the loudness at the ears. This type of asymmetry changes the firing rate at one ear, relative to the other. Both of these are used together in order to correctly localize an auditory stimulus.

1.4 Research Goals

In the current thesis, two studies set out to characterize the neuroelectric correlates of auditory object boundaries. More specifically, we investigated *first-order* and *second-order* auditory boundaries. Previously reported correlates of auditory object processing in the auditory evoked potential have prominently focused on responses to stimuli that onset abruptly relative to a quiet background. From the perspective of auditory scene analysis, such first-order boundaries are less interesting than other types of boundaries. Therefore, Experiment 1 in Chapter 2 sought to differentiate responses to stimulus onsets (with abrupt changes in acoustic energy) and responses to stimulus pitch transitions (containing no energy increment). Previous research showed that when lateralized stimuli were presented, second-order motion-defined boundaries exhibited a robust evoked potential at lateral electrodes whereas first-order boundaries did not (Butcher *et al.*, 2011). Experiment 2 in Chapter 2 sought to further investigate responses when stimuli were lateralized. Specifically, this experiment considered whether the responses to lateralize pitch boundaries exhibited similar lateralization of waveforms as responses to motion boundaries. Analysis of the data focused on the grand-averaged ERP waveform differences as well as time-frequency measures. Specifically, time-frequency analysis focused on inter-trial phase coherence. In order to make claims as to first-order

and second-order boundaries in general terms, the goal of Chapter 3 was to investigate whether the differences between first-order and second-order processing of pitch boundaries were also evident in response to second-order motion boundaries. Therefore, in Chapter 3 we investigated the responses to first-order and second-order auditory motion boundaries. In addition, Chapter 3 allowed for replication of the Butcher et al. (2011) study with first-order and second-order boundaries that were more clearly defined and evaluated. More specific goals for both Chapter 2 and Chapter 3 are presented at the beginning of these chapters.

Chapter 2: Brain Dynamics Encode the Spectrotemporal Boundaries of Auditory Objects

2.1 Preface

The fundamentals of scene analysis were introduced to me only a few years ago in the context of the visual scene. The ability for one to effortlessly parse a visual scene into its objects is an incredible task. It is one task that we are constantly engaged in, whether we realize it or not. What is it about particular boundaries that allow us to segregate one object from another or from a background? How is it that we can pick out particular characteristics of an object – its redness and truck-ness – and put them together to form a coherent object – a fire truck? Furthermore, we can take that object and ensure that it does not become fused with its background. It may seem intuitive when speaking in terms of the visual scene since these are objects that people easily interact with, touch, and move. An even more incredible task, and one that seems more chaotic, is that of auditory scene analysis. The spectrotemporal properties of a sound object or 'stream' are constantly bombarding the basilar membrane. The brain can take these properties from many incoming signals and segregate them into their respective streams and even focus on one over another. I was immensely interested in how this task is accomplished and the mechanisms the brain uses in order to differentiate between particular aspects of an acoustic stream.

For this reason, I chose to focus on the neuroelectic correlates of boundaries that are present in acoustic environments. In addition, I wanted these environments to have dynamic aspects. In the experiment in this chapter, I focused on responses to first-order

and second-order boundaries that were defined by spectrotemporal discontinuities in pitch when the stimuli are presented in front of and to the side of a listener.

2.2 Introduction

We effortlessly parse complex scenes into objects and background. In vision science, a substantial effort has been made to understand how the visual system is able to group surfaces and edges together into objects. These mechanisms are believed to be modular and hierarchical so that different aspects of visual input can be handled by specialized structures (Marr, 1982; Zeki *et al.*, 1991; Goodale & Milner, 1992). Despite much progress in vision science, the mechanisms by which the auditory scene is analyzed for object boundaries remain unclear. Anatomical and functional evidence regarding the specificity of 'what' and 'where' *auditory* pathways suggests that the general process might be broadly analogous to that of the visual system (Rauschecker & Tian, 2000; Alain *et al.*, 2001a; Kubovy & Van Valkenburg, 2001; Tian *et al.*, 2001). The goal of the two experiments in this study was to identify neural signals associated with boundary or edge detection in the auditory system.

In vision, segmentation of objects and background can occur in response to first-order boundaries such as discontinuities in luminosity or brightness. Higher-order boundaries can be defined in terms of discontinuities in color (Shapley, 1990), texture (Mareschal & Baker, 1998), or motion (Anderson, 1997). While first-order boundaries appear to be processed in the initial feedforward sweep of information from the medial geniculate nucleus to V1 and onward, higher-order boundaries have been considered to require feedback processing. This feedback processing entails information being

processed in visual areas such as V2 and projecting back to V1 in order to form the complete perception (Lamme *et al.*, 1994; Lamme & Roelfsema, 2000; Lee & Nguyen, 2001). In order to investigate first-order and higher-order boundaries in the auditory scene it is important to note that while the visual scene is spatiotopic, the auditory scene is tonotopic. Therefore, auditory boundaries are spectrotemporal in nature (Griffiths & Warren, 2002; 2004) and can be distinguished by discontinuities in the temporal evolution of features. In this sense, first-order auditory boundaries are increments or decrements of sound energy over time and higher-order auditory boundaries are spectrotemporal discontinuities such as those in pitch and interaural timing. For example, auditory boundaries can be defined and studied by temporal discontinuities in dichotic pitch (Ungan *et al.*, 1989; Johnson *et al.*, 2003), mistuned harmonics (Alain *et al.*, 2001a; Alain *et al.*, 2001b; Alain *et al.*, 2002), energy-matched pitch (Krumbholz *et al.*, 2003; Seither-Preisler *et al.*, 2004; Seither-Preisler *et al.*, 2006b), motion (Butcher *et al.*, 2011), organized and disorganized sequences of tones (Chait *et al.*, 2007a; Chait *et al.*, 2008), and interaural correlation (Chait *et al.*, 2007b). In each of these examples, however, the distinction is not made between the stimulus features that support boundary detection and the functional anatomy that performs boundary detection. For example, it is not known how the functional anatomy of objects in vision corresponds to the functional anatomy of objects in audition. Non-analogous structures might perform analogous computational tasks. In other words, while the pathways from primary visual cortex that visual object information takes are relatively well understood, the analogous pathways for auditory object information from primary auditory cortex are much less clear.

Electroencephalography (EEG) and magnetoencephalography (MEG) have been useful in identifying brain responses to boundaries in the auditory scene. Robust deflections of the associated event-related potential (ERP) and the analogous event-related field (ERF) in MEG are triggered by the onsets of acoustic stimuli (Hillyard & Picton, 1978; Shahin *et al.*, 2007). Well-known components of the ERP/ERF are modulated by both sensory and cognitive factors. For example, the N1 component of the ERP is modulated by attention (Hillyard *et al.*, 1973; Näätänen & Picton, 1987) and the P300 components reflects not only cognitive factors, such as novelty of a stimulus, but also psychophysiological context, such as arousal (Polich & Kok, 1995). Other components of the event-related waveforms offer specific clues about neural mechanisms that represent boundaries in the environment. However, sound onsets typically have energy transients - increments or decrements in total acoustic power. Thus, ERP/ERF responses to first-order auditory boundaries are well studied, but offer little opportunity to distinguish low-level sensory responses from high-order boundary-detection processes.

Alain and colleagues (Alain & Arnott, 2000; Alain *et al.*, 2002; Dyson & Alain, 2004; McDonald & Alain, 2005; Snyder *et al.*, 2006) identified a component of the auditory ERP that was associated with the registration of discrete objects in the auditory scene and termed this deflection the object-related negativity (ORN). Similarly, Butcher et al. (2011) identified an ERP component that was evident when auditory objects could be segmented by discontinuous motion trajectory in a virtual auditory scene. This deflection occurred on the side contralateral to the sound location and they accordingly termed this deflection the lateralized object-related negativity (LORN). Second-order boundaries can also be investigating using Iterated Rippled Noise (IRN), which applies a

delay and add algorithm to broadband noise. Transitions from broadband noise to the IRN have the same overall energy, but there is a perceptual change in pitch. In a set of investigations using iterated rippled noise (IRN), Krumbholz et al. (2003) reported an MEG response evoked by the transition from background noise to a discrete pitch with no acoustic energy change. They termed this the pitch onset response (POR). An important design feature shared by these lines of research is that perceptual objects were defined by spectrotemporal boundaries rather than energy boundaries. This allowed a distinction to be made between energy-related and other boundary-related responses. For example, the POR is at least partially distinct from a more general energy-onset response (Krumbholz *et al.*, 2003; Seither-Preisler *et al.*, 2004; Seither-Preisler *et al.*, 2006a).

Previous work on the neural correlates of auditory boundary representation has emphasized ERP/ERFs. The EEG/MEG signal is composed of power in several frequency bands, but the ERP/ERF tends to be dominated by low-frequency power in the theta (4-8Hz) and alpha (8-12Hz) bands. Time-frequency decomposition allows for investigation of changes in signal composition at many frequency bands simultaneously. In addition, ERP/ERF waveforms are averages of many trials and only reveal signals with high temporal coherence. However, analysis of the spectrotemporal decomposition of single trials can reveal signals that are not strictly phase-locked to stimuli (Ponjavic-Conte *et al.*, 2013). In Experiment 1 below, we contrast boundaries defined by energy (first-order) with boundaries defined by pitch (second-order), with emphasis on characteristic differences in the time-frequency decomposition of the EEG. Our theory for Experiment 1 was that first-order auditory boundaries and higher-order boundaries would be processed by different neural mechanisms. Our prediction was that higher-order

boundaries would require feedback processing and would therefore exhibit ERP differences in the 100-150ms latency window. Meaning that first-order boundaries would be processed earlier than second-order boundaries. In Experiment 2, we considered how lateralization of the stimuli would affect responses to these boundaries.

2.3 Experiment One Methods

2.3.1 Participants

Nineteen students from the University of Lethbridge participated in Experiment 1 for course credit. The nature of the experiment was explained to the participants. Written informed consent was completed and the participants reported no hearing or neurological problems. Procedures were in accordance with the Declaration of Helsinki and approved by the University of Lethbridge Human Subjects Review Committee. Trials contaminated with eye artifact (>120 μV) were rejected and data from subjects with less than 50% of original trials were not analyzed, leaving fifteen participants (five male; mean age: 20.1; one left-handed).

2.3.2 Stimuli

The goal of this experiment was to measure differences in brain responses to different kinds of spectrotemporal boundaries. We refer to boundaries as ‘first order’ if they entailed a broadband energy increment or decrement relative to background noise (Fig. 2.1a). These first-order stimuli always contained an abrupt discontinuity in energy at all frequency bands. Therefore, these first-order stimuli always contained an abrupt discontinuity in energy at all frequency bands. We refer to boundaries as ‘second order’ if

they entailed spectrotemporal discontinuities, but the total acoustic energy of the stimulus remained constant (Fig. 2.1b). These second-order boundaries also contained some spectral discontinuity in amplitude at particular frequencies. However, the overall energy of the stimulus remained constant through time. Meaning that while global acoustic energy remains constant, there are local regions that experience increments while other local regions experience decrements. We used iterated rippled noise (IRN) to create second order boundaries. The IRN is produced using a delay-and-add algorithm developed by Yost (1996). The pitch of the IRN is determined by the reciprocal of the delay time and the salience is determined in part by the number of iterations of the delay. For the first experiment our IRN stimuli had 4 iterations of a time delay of 40 samples, given a 32000Hz sample rate, resulting in a 1.25ms time delay. The perceived pitch of these IRN stimuli was 800Hz. IRN stimuli were scaled so that the root-mean-square (RMS) amplitude matched that of the temporally adjacent broadband noise over the entire spectrum. We also considered that the transitions between noise and IRN could be marked by changes in loudness despite constant RMS.

We applied Moore's time-varying loudness model (Glasberg & Moore, 2002) and confirmed that the silence-to-IRN stimuli did exhibit loudness modulation (Fig. 2.1c), whereas the noise-to-IRN boundary did not (Fig. 2.1d). This measure allowed us to ensure that second-order boundaries did not have an overall change in loudness (a first-order boundary). It is important to note that, by definition, boundaries between broadband noise and IRN feature energy increments at some frequencies and decrements at others, which can be seen in Figure 2.1e and 2.1f for first-order and second-order boundaries,

respectively. Figure 2.1 also shows the spectral composition of the IRN stimulus (Fig. 2.1g) and the white noise (Fig. 2.1h).

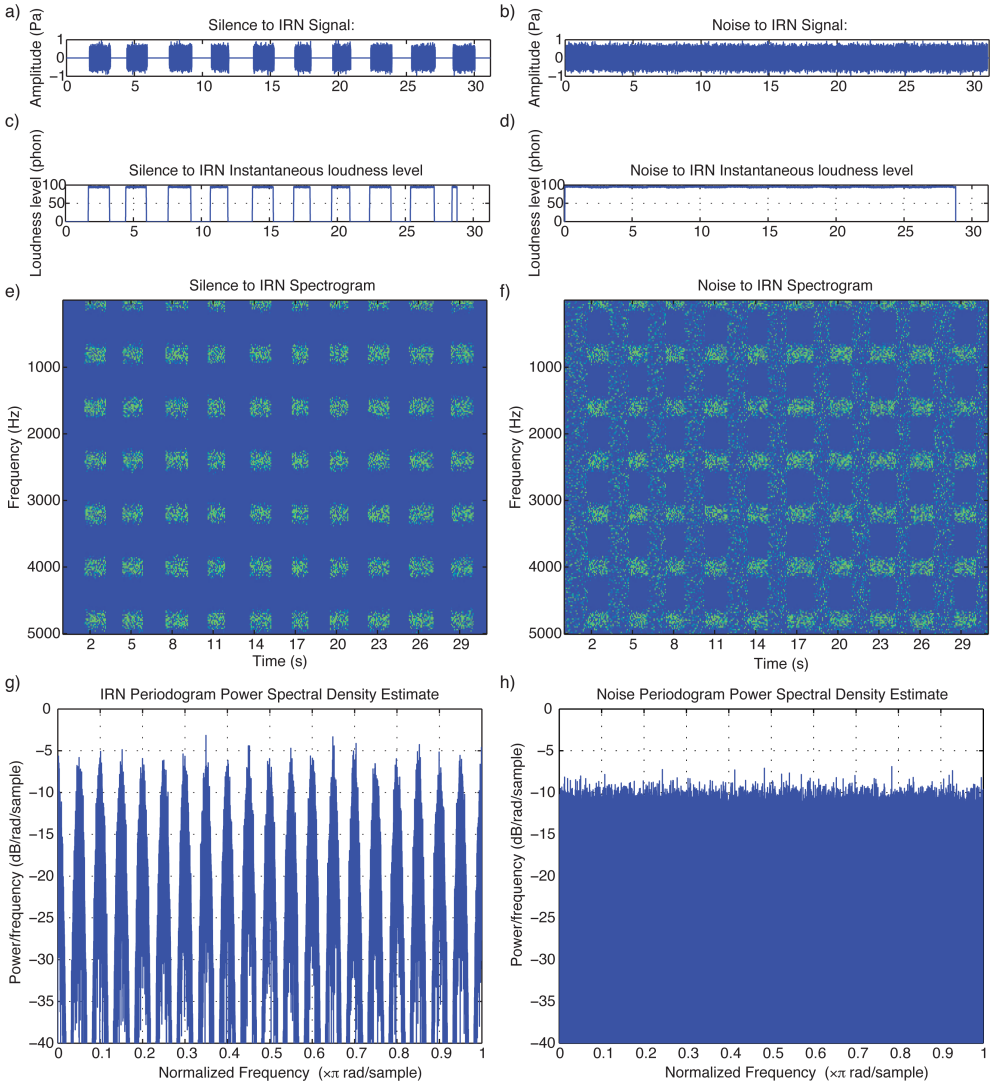
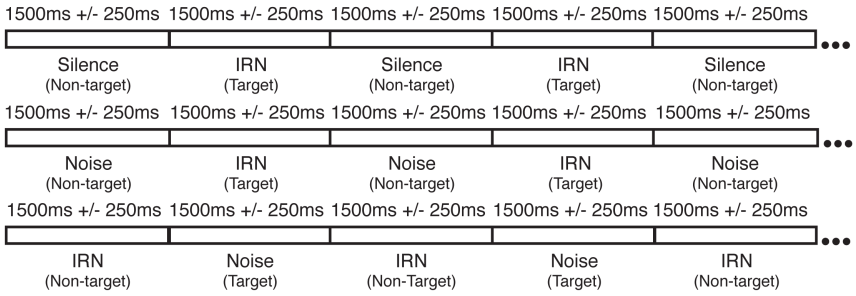


Figure 2.1: (a) The acoustic waveform itself indicates the energy envelope of the silence-to-IRN and IRN-to-silence boundaries. (b) The noise-to-IRN and IRN-to-noise boundaries are not visible in the waveform because the total energy remains constant. (c) The Glasberg and Moore model of loudness applied to the silence-to-IRN stimulus. (d) The Glasberg and Moore model of loudness applied to the noise-to-IRN stimulus. (e) The spectrogram of the silence-to-IRN stimulus, which shows the spectral composition of both broadband noise and IRN. (f) The spectrogram of the noise-to-IRN stimulus, which shows the spectral composition of IRN. (g) The periodogram of the IRN, which has linear spacing in the frequency domain. (h) The periodogram of the broadband (white) noise, which is completely random in the frequency domain.

Experiment 1 had three stimulus conditions. These varied in the type of spectrotemporal discontinuity used to define a target sound. In the silence-to-IRN condition, IRN stimuli were interleaved with silent intervals. All responses were made by pressing the spacebar via a computer keyboard. Participants responded when they detected the onsets of IRN (i.e. IRN was the target). In the noise-to-IRN and the IRN-to-noise conditions the IRN stimulus alternated with broadband noise. Participants responded to IRN onsets in the noise-to-IRN condition (i.e. IRN was the target) and they responded to noise onsets in the IRN-to-noise condition (i.e. noise was the target). Stimuli were presented through a speaker (Mackie HR-624mk2 studio monitor) directly in front of the participant. Twenty blocks of each condition were presented in pseudorandom order. Blocks consisted of repeating cycles of non-target sounds alternating with target sounds. Boundaries were jittered so that the duration of each



stimulus was between 1250ms and 1750ms (Fig. 2.2). Each block took an average 31.5 seconds and included 10 target presentations. Participants took self-paced rest breaks between blocks.

Figure 2.2: Stimulus presentation for Experiment 1. Ten cycles of 1250ms to 1750ms of non-target stimuli followed by 1250ms to 1750ms of target stimuli and ending with 1250ms to 1750ms of non-target.

2.3.3 Setup

Participants sat in a sound attenuated room with anechoic wall covering. Background noise floor was approximately 60 dB of broadband noise, primarily from HVAC airflow. A video monitor was placed in front of the participants, slightly below the front speaker, and participants were instructed to keep their gaze on a fixation dot in the middle of the screen.

2.3.4 Recording and Analysis

The EEG was recorded at 500Hz using 128 silver/silver chloride electrodes in a plastic net and EGI's Netstation acquisition software (Electrical Geodesics, Inc., Eugene, OR, USA). For a diagram of the international EEG electrode placement locations, see Appendix A. Impedances were maintained below 100kΩ. Further analysis of the EEG was done with BESA software (Megis Software, Gräfelfing, Germany) and customized MATLAB (MATLAB 2010b, The Mathworks Inc., 2010, Natick, Massachusetts, USA) tools based on the EEGLab toolbox (Delorme & Makeig, 2004) and the Fieldtrip toolbox (Oostenveld *et al.*, 2011). EEG recording was visually inspected for bad channels and no more than eight electrodes were replaced with interpolated signal using spherical spline interpolation (Baramidze *et al.*, 2006). Digital filters were applied to the EEG data: high-pass (0.5Hz, 6dB/octave, first-order) forward-phase and low-pass (30Hz, 24dB/octave, fourth order) zero-phase Butterworth filters. Trials contaminated with eye artifact (>120 μV) were rejected from the average. The averaged ERPs were re-referenced and interpolated to a 10-10 reference-free montage (Luu & Ferree, 2005).

2.3.5 Behavioral Data

Average reaction times for Experiment 1 were computed for each condition in MATLAB. Only trials on which the participant correctly identified the target were included in these averages. The data were analyzed using a one-way repeated-measures analysis of variance (ANOVA) with three levels of the factor boundary type (silence-to-IRN, noise-to-IRN, and IRN-to-noise).

2.3.6 Event-Related Potential Waveforms

For Experiment 1 we computed the ERP waveforms over 1000ms epochs (timespan), time-locked to auditory boundaries (both onsets and offsets). The waveforms were grand-averaged for each condition with an epoch of -200ms to 800ms using a baseline of 200ms to 0ms. Only trials in which the participant correctly detected the target were included in the average. In addition to computing ERPs for the three types of target onsets, we also computed ERPs time-locked to the target offsets. This enabled us to consider differences due solely to task parameters. For example, onsets of IRN targets in the noise-to-IRN condition were the same physical stimulus as offsets of the noise target in the IRN-to-noise condition, but required different behaviour from the participant according to the task. Scalp topographies of the components in the grand-averaged waveforms were visualized with isopotential maps. The amplitude of the P90 and N1 were analyzed by computing the mean amplitude under a 20ms latency window determined from the grand average. The latency of the N1 was analyzed using Brain Electrical Source Analysis (BESA) peak selection and a one-way repeated-measures ANOVA with three levels of the factor boundary type was carried out.

2.3.7 Time-frequency Analysis

We investigated the inter-trial phase coherence (ITPC) and time-spectral evolution (TSE) of power from the raw data in order to investigate responses at multiple frequency ranges. The signal was transformed into time-frequency space through complex demodulation as described by Hoechstetter et al. (2004).

(1) $$Z_{k,t,f} = A_{k,t,f} \cdot e^{i\theta_{k,t,f}}$$

(2) $$ITPC_{t,f} = \left| \frac{1}{N} \sum_{k}^{N} e^{i\theta_{k,t,f}} \right|$$

(3) $$B_f = \frac{1}{n_t} \sum_{t_{prestimulus}}^{t_0} \frac{1}{N} \sum_{k}^{N} \left| Z_{k,t,f} \right|^2$$

(4) $$TP_{t,f} = \frac{\frac{1}{N} \sum_{k}^{N} \left| Z_{k,t,f} \right|^2}{B_f}$$

(5) $$EP_{t,f} = \frac{\left| \frac{1}{N} \sum_{k}^{N} Z_{k,t,f} \right|^2}{B_f}$$

(6) $$IP_{t,f} = TP_{t,f} - EP_{t,f}$$

where $A_{k,t,f}$ is the amplitude of the complex valued result ($Z_{k,t,f}$) of the complex demodulation for trial k, frequency f, and time *t*; $\theta_{k,t,f}$ is the phase; N is the number of trials; B_f is the baseline power; n_t is the number of time bins; $TP_{t,f}$ is the total power (percent change from baseline); $EP_{t,f}$ is the percent change in power from baseline that is evoked (i.e. phase-locked); and IPt,f is the percent change in power from baseline that is induced (i.e. non-phase-locked).

ITPC (Equation 2) measures how well the phases of signals match over successive trials. This means that the responses are occurring with high coherence. TSE (Equation 4) is the percent change of power from the baseline and shows the time-locked components (both 'induced' and 'evoked'). Evoked power (EP) (Equation 5) is the component of the EEG signal that is phase-consistent (i.e. power that is phase locked) across trials. We computed EP by averaging the complex value output of the complex demodulation across trials. Induced power (IP) (Equation 6) is the component of the EEG signal that is non phase-locked.

For Experiment 1, the frequency bins were incremented by 2Hz between 4 and 60Hz and the time bins were incremented by 25ms from -200 to 800ms. Plots for ITPC and TSE power were computed for all three target boundary transitions. The statistical significance of differences between experimental conditions was assessed using non-parametric, Monte Carlo permutation tests. These tests were used because the distribution of the data is not known. The permutation was repeated 50000 times for each participant. To control for multiple comparisons we applied a false-discovery rate (FDR) correction method (Benjamini & Hochberg, 1995). Significant effects of condition were found using an FDR corrected alpha value corresponding to $\alpha=0.05$.

2.4 Experiment One Results

2.4.1 Behavior

For Experiment 1, average reaction times were 528.63ms for first-order silence-to-IRN boundaries, 509.31ms for second-order noise-to-IRN boundaries, and 520.16ms for second-order IRN-to-noise boundaries. The effect of boundary type on reaction time

did not reach significance ($F_{2,28}$=1.727; p=0.196). The hit rate (percentage of correct responses) was 88.9% for first-order boundaries, 92.9% for second-order noise-to-IRN boundaries, and 94.4% for second-order IRN-to-noise boundaries. The effect of boundary type on hit rate was marginally significant ($F_{2,28}$=2.705; p=0.084).

2.4.2 Event-Related Potentials & Isopotential Voltage Maps

ERP data for all boundary types in Experiment 1 are shown in Figure 2.3. First-order boundaries (silence-to-IRN) triggered ERP responses with P90, N1, and P2 deflections (Fig. 2.3a). Similarly, transitions from IRN to noise evoked P90, N1, and P2 deflections (Fig. 2.3c & e). These deflections were observed regardless of whether the IRN to noise transition signified a target onset or offset. Transitions from noise to IRN evoked notably different ERP responses. The P90 response observed for first-order boundaries and IRN-to-noise transitions was absent for noise-to-IRN transitions. Instead, the earliest ERP peak in this condition was the N1 (Fig. 2.3b & f). This distinction held regardless of whether the transition signified a target onset or offset. Note that a noise-to-IRN offset and the IRN-to-noise onset are the same physical stimulus, but differ in task relevance. Likewise, IRN-to-noise offsets and noise-to-IRN onsets are the same physical stimulus, but differ in task relevance. The ERP peaks in every condition were most prominent at fronto-central electrode Fz. At this electrode, the N1 peaks of the averaged target ERPs occurred at 126ms for first-order boundaries and 136ms for both types of second-order boundaries. The N1 deflections appeared smaller in amplitude for first-order boundaries compared to both types of second-order boundaries. However, the effect of boundary type on N1 amplitude did not reach significance ($F_{2,28}$=1.979; p=0.157).

There was an effect of boundary type on N1 latency when comparing all boundary types ($F_{2,28}$=9.083; p=0.001) indicating that the N1 peak in response to first-order boundaries occurred earlier than the N1 for second-order boundaries. Post-hoc t-test analysis indicated that first-order silence-to-IRN N1 occurred earlier than second-order noise-to-IRN boundaries (t=3.465; p=0.004) and earlier than second-order IRN-to-noise boundaries (t=-3.862; p=0.002). However, there was no significant effect of boundary type on the N1 latency between second-order conditions (t=-1.517; p=0.151).

For transitions that triggered a P90 response (i.e. silence-to-IRN and IRN-to-noise), isopotential voltage maps at 90ms indicate foci over the right frontal scalp. At the latency of the N1 peak, isopotential voltage maps indicate a negative focus over frontocentral scalp with polarity reversals at posterior-lateral sites, suggesting symmetric bilateral auditory cortex generators of the N1 for both first- and second-order boundaries. At the latency of the P2 peak, responses were maximal over midline frontal scalp for both first- and second-order boundaries.

During the silence-to-IRN offset transition, the amplitude of the evoked potential was greatly attenuated and isopotential maps exhibited little similarity to the other boundary types (Fig. 3d).

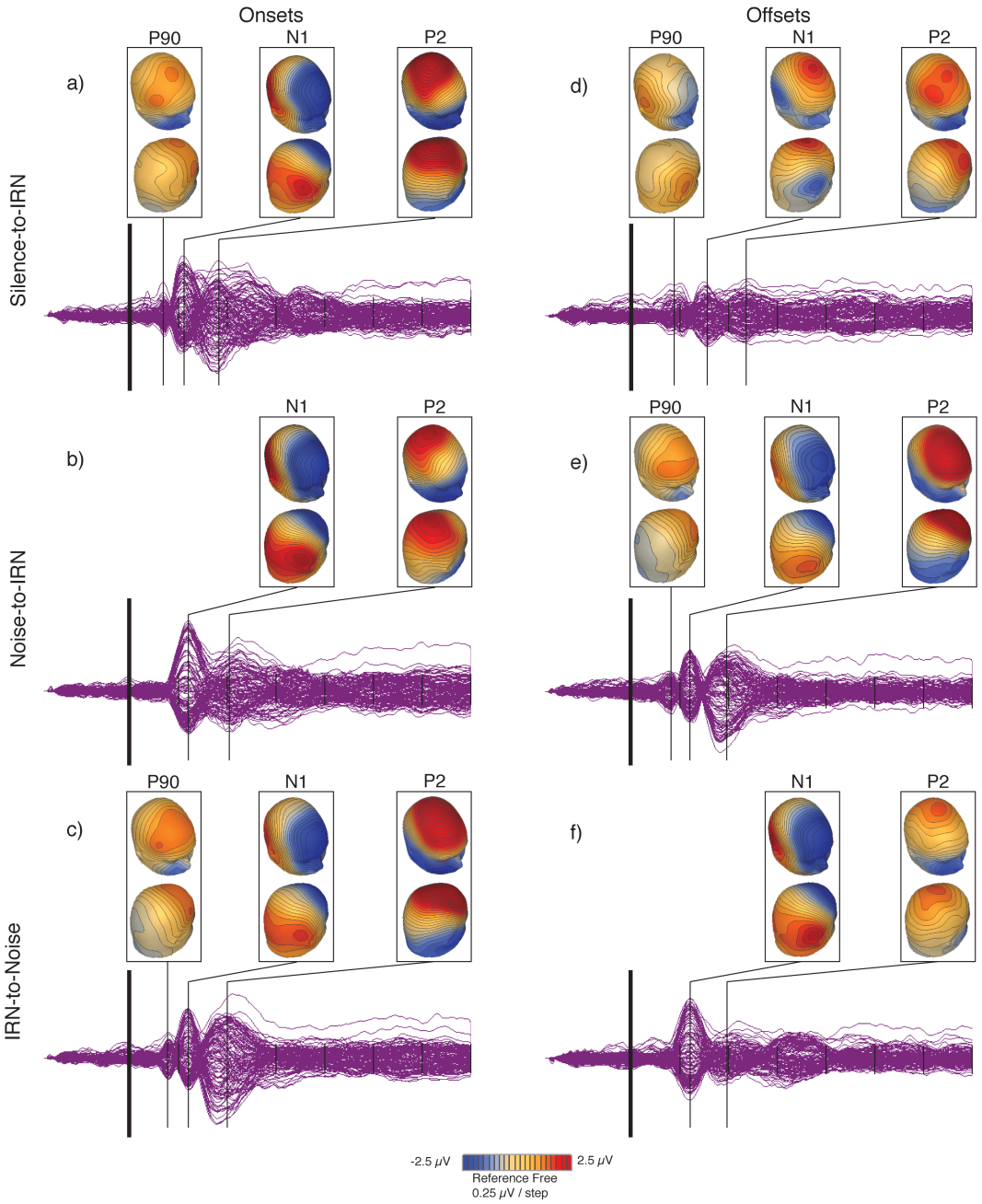


Figure 2.3: ERP waveforms and isopotential voltage maps to target onsets (left column) and offsets (right column). Isopotential maps are shown for prominent peaks in the ERP waveforms. (a) First-order silence-to-IRN transition; IRN onset. (b) Second-order noise-to-IRN transition; IRN onset. (c) Second-order IRN-to-noise transition; noise onset. (d) First-order silence-to-IRN transition; silence onset. (e) Second-order noise-to-IRN transition; noise onset. (f) Second-order IRN-to-noise transition; IRN onset.

2.4.3 Time-Frequency Analysis

2.4.3.1 Theta and Alpha Bands

Time-frequency decomposition for Experiment 1 is shown in Figure 2.4 for electrode Fz. All boundaries triggered phase-locked power in the theta and alpha bands around the N1 latency, as evidence by inter-trial phase coherence (ITPC) (Fig. 2.4; column 1) and evoked power (Fig. 2.4; column 3). The results of the permutation test, comparing EEG data associated with first- and second-order boundaries as well as comparing the two kinds of second-order boundaries are displayed in Figure 2.4, rows 4-6. These data are masked to show only time-frequency bins with p-values less than the false-discovery rate (FDR). Second-order noise-to-IRN boundaries exhibited less total evoked power in the theta and alpha bands compared to first-order boundaries, particularly at the frequency of 8Hz and latency of the N1 component. The second-order IRN-to-noise boundaries exhibit similar evoked power in the theta and alpha bands compared to first-order boundaries, but it occurs slightly later in time. Only two time-frequency bins passed the FDR threshold for induced power in the first-order silence-to-IRN compared to second-order IRN-to-noise boundaries. Two time-frequency bins passed FDR threshold for less total TSE power for second-order noise-to-IRN boundaries compared to first-order boundaries as well as for second-order noise-to-IRN compared to second-order IRN-to-noise boundaries. However, these overlapped with significant regions of less evoked power.

There was robust induced power at approximately 700-800ms in the 20-30Hz frequency range. Because the time-frequency plots in figure 2.4 depict target transitions, this increase in power is likely related to motor response.

When non-target transitions were investigated, there was an attenuated response in the silence-to-IRN offset compared to both second-order boundary offsets, which can be seen in the ERP waveforms. There were no other significant difference between offset conditions and, therefore, time-frequency plots are not shown.

2.4.3.2 Gamma Band

Boundary types differed notably in an early evoked transient signal above the theta and alpha bands. First-order onset boundaries in the silence-to-IRN conditions triggered an early phase-locked transient response mainly in the gamma band (Fig. 2.4; row 1, column 1). This gamma signal was absent in both of the second-order boundary conditions (Fig. 2.4; rows 2-3, column 1). The gamma difference between first- and both second-order boundaries was also revealed in the permutation test (Fig. 2.4; rows 4-5). The differential gamma signal associated with first-order boundaries also appears in the plots of total TSE power. However, this is probably due to the contribution of evoked gamma power to total power. There was one time-frequency bin that passed FDR in the gamma band for second-order noise-to-IRN compared to second-order IRN-to-noise boundaries (Fig. 2.4; row 6).

The observation of this transient gamma response to first-order boundaries led us to reconsider the frequency composition of the grand averaged ERP waveforms shown in figure 2.3. When the grand-averaged ERP waveforms were band-pass filtered at 28-48Hz (Pantev *et al.*, 1991), there was also evoked gamma at electrodes close to Fz.

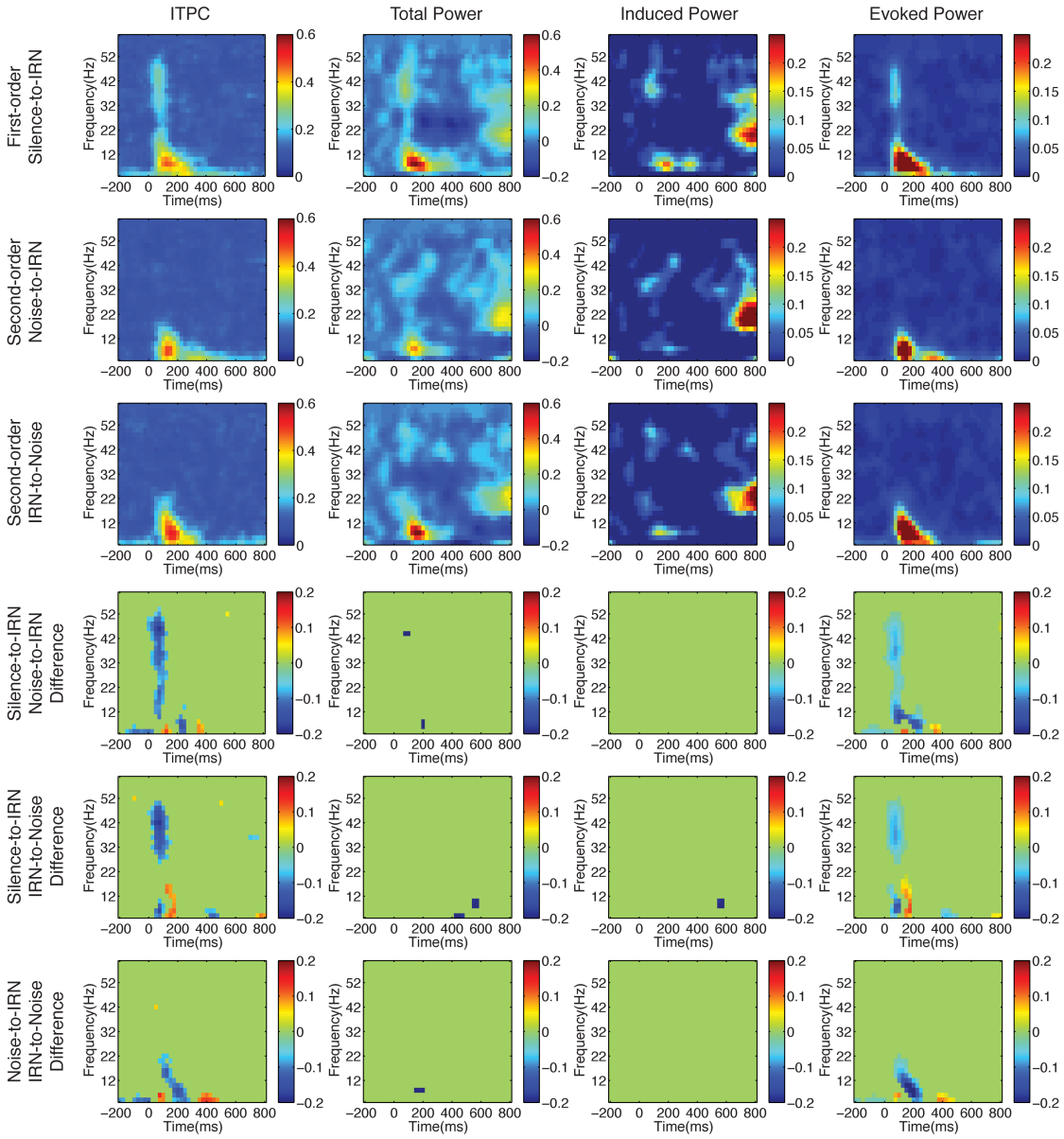


Figure 2.4: Time-frequency analysis for target onsets, including inter-trial phase coherence (ITPC), total power, induced power, and evoked power, respectively in columns. Row 1: First-order boundaries; Row 2: Second-order noise-to-IRN boundaries; Row 3: Second-order IRN-to-Noise boundaries. Rows 4-6 display the difference between the above conditions, masked to preserve false-discovery rate. The value of each bin is displayed only if that bin's p-value in the permutation test was less than the FDR cut-off. Bins that did not pass this FDR test are set to green.

2.5 Experiment One Discussion

Experiment 1 contrasted target sounds defined by various kinds of spectrotemporal discontinuities in the auditory scene. These spectrotemporal discontinuities could be primarily defined by energy increment (silence-to-IRN) or decrement (IRN-to-silence), or they could be defined by discontinuities at specific bands of the spectrum (noise-to-IRN and IRN-to-noise). The results of Experiment 1 showed that different kinds of boundaries are encoded via different mechanisms in the auditory system.

First, we consider whether differences in perceptual salience drove the observed EEG/ERP differences, as it seems possible that the transition from silence to IRN would be substantially more salient than a transition from noise to IRN. This is unlikely because reaction times did not differ significantly and hit rate was near ceiling for all boundary types. From the initial theory, that first-order boundaries would be processed earlier than second-order boundaries, a prediction that could be made is that responses to first-order boundaries would be faster and more accurate. However, neither reaction times nor hit rate to first-order boundaries were significant compared to second-order boundaries. Indicating that the responses to second-order boundaries are evidently not merely less robust versions of first-order responses - both performance and electrophysiological data suggest that the second-order boundaries are at least equally perceptually salient. This is consistent with previous observations of the pitch onset response (POR) (Krumbholz *et al.*, 2003; Seither-Preisler *et al.*, 2004) and lateralized object-related negativity (LORN) (Butcher *et al.*, 2011).

The ERP data suggest that first- and second-order auditory boundaries engage different neuronal mechanisms within the first 200 milliseconds after an auditory boundary occurs. In all of the stimulus conditions investigated here, discontinuities in the acoustic envelope that defined target onsets elicited robust ERP waveforms with prominent N1 potentials over the frontal midline, particularly at electrode Fz. Interestingly, the second-order IRN-to-noise boundary elicited a robust waveform when the transition was a target as well as when it was a non-target. This is not consistent with previous POR work by Krumbholz et al. (2003) and Lütkenhöner et al. (2011). Both of these prior studies found no discernible MEG response due to the pitch to noise transition. It could be that the evoked potential found in the current study is due to EEG being more sensitive to this transition. Alternatively, it could be that the evoked response is not specific for pitch onsets and reflects a more general response to an object boundary. Waveforms evoked by target offsets were smaller in amplitude, which is consistent with previous work (Pfefferbaum *et al.*, 1971; Hillyard & Picton, 1978; Hari *et al.*, 1987). The latency of the N1 peak associated with second-order boundaries was later than that of a classic auditory evoked potential. This is consistent with latency of the N1 reported in MEG work by Chait et al. (2006) and Lütkenhöner et al. (2011) for the POR as well as transitions from a random sequence of tones to a constant tone (Chait *et al.*, 2007a; Chait *et al.*, 2008).

The current experiment revealed two differences between various spectrotemporal boundaries in the auditory scene, which we discuss in turn below: First, second-order noise-to-IRN boundaries conspicuously lacked the early P90 response that was evident in

both first-order boundaries and second-order IRN-to-noise boundaries. Second, first-order boundaries alone evoked a robust, phase-locked gamma signal.

Second-order boundaries defined by transitions from noise to IRN are spectrotemporally complex stimuli. These consist of an instantaneous increment in energy at some frequency bands and a simultaneous decrement of energy at other bands (Fig. 2.1f). Thus, although the total acoustic energy in the scene remains constant across a second-order boundary, the energy impinging on any one region of the basilar membrane may increase or decrease, depending on the frequency to which that region is tuned. Conversely, the second-order boundaries defined by transitions from IRN to noise are characterized by the onset of energy at some frequency bands and a simultaneous decrement of energy at those frequencies that are components of the IRN tone. This may explain why IRN-to-noise boundaries resemble first-order silence-to-IRN boundaries with respect to the P90 peak: both contain energy increments. The noise-to-IRN boundaries contained no such energy increments; instead these contained only energy decrements and incrementally greater energy at IRN components. We therefore speculate that the P90 component preceding the N1, which is evoked for silence-to-IRN and IRN-to-noise boundaries is related to spectral increments that are missing from the noise-to-IRN boundary.

When further exploring the differences between first- and second-order boundaries in the auditory scene, we considered several measures of brain oscillatory dynamics: inter-trial phase coherence (ITPC), total time spectral evolution (TSE) of power and phase-locked 'evoked' power (EP). All auditory boundaries triggered an event-related increase in power in the alpha and theta frequency bands at roughly the

latency of the N1 ERP component. In each case this power was evoked. That is, it exhibited substantial ITPC and was, therefore, probably associated with the N1-P2 complex in the ERP waveform. First-order boundaries differed from second-order boundaries in that these also evoked an early transient high-frequency response mainly in the gamma band. Second-order boundaries exhibited no such gamma response. This gamma signal was also not evoked by first-order boundary offsets. It is therefore of some interest to consider whether this transient gamma signal might have functional significance. This transient gamma has previously been reported in auditory ERP waveforms that have been filtered to show the 28-48Hz frequency range (Pantev *et al.*, 1991; Pantev & Elbert, 1994). These studies also indicated that the gamma band response is strongly locked in phase during the 20-120ms latency.

Experiment 1 demonstrated that auditory activity is evident not only in the theta and alpha bands, but also in the gamma frequency band, which may be hidden in the ERP waveform by the larger theta and alpha response. Importantly, the gamma band signal was evoked. That is to say, it was tightly time-locked to the boundary and thus it exhibited a high degree of inter-trial phase coherence. The earliest auditory-evoked electrical responses typically measured at the scalp are the brainstem evoked responses and mid-latency components. These are relatively high frequency and might give rise to the early transient gamma we observe for first-order onsets. However, the gamma activity we observed occurred somewhat later, at 75ms. Gamma band signals in monkey visual cortex have been associated with neuronal mechanisms of feature integration (von der Malsburg & Schneider, 1986; Singer, 1999) and the fine temporal structure of gamma oscillations is believed to be a mechanism of attentional feature selectivity (Fries *et al.*,

2002; Fries, 2005). Induced gamma in the human EEG has been associated with a variety of visual events that require global integrations of spatially disparate information over the visual scene (Tallon-Baudry & Bertrand, 1999; Doesburg *et al.*, 2005). Furthermore, recent work showed that gamma activity in visual cortex is significantly increased when communication between neurons in primary visual cortex and extrastriate (particularly V4) is feedforward (V1 to V4) or bottom-up (Bosman *et al.*, 2012). The IRN stimulus employed here is spectrally complex, consisting of energy at distinct pitch bands separated by bands at which there is no acoustic energy as described above. Even so, the IRN stimulus is perceived as a single coherent sound object. Since the auditory system is tonotopic rather than spatiotopic, the IRN stimulus is notionally similar to an array of spatially distinct elements that are integrated into a gestalt. One possibility is that gamma activity associated with first-order boundaries reflects a bottom-up or feed-forward signal that triggers integration across the neuronal representations of the distinct frequency components of the IRN sound. However, the auditory evoked gamma signal that we observed in this experiment falls between 35 and 45 Hz. It is therefore not clear whether this signal shares functional relevance with the somewhat higher frequency activity typically observed in visual cortex of monkeys. Any hypothesis linking the two necessarily remains purely speculative and awaits further investigation.

The gamma observed in this experiment exhibits inter-trial phase coherence (i.e. evoked), which is of particular interest because many studies report ‘induced‘ rather than ‘evoked’ gamma. This may be due to the high inter-trial temporal precision we were able to achieve in scheduling playback in our audio software and hardware design (+/- 1ms). Temporal jitter of only a few milliseconds in an evoked gamma signal effectively shunts

power from evoked to induced power (David *et al.*, 2006). The perceptual consequences of evoked gamma in this experiment remain unclear because listeners performed at ceiling in the detection task. However, Pantev et al. (1994) suggested that cortical analysis of auditory stimulus features are represented by slow auditory evoked activity, whereas the gamma band response represents more general perception mechanisms. Busch et al. (2006) found that bottom-up or feedforward factors modulated phase locking, whereas top-down or feedback factors modulated the power of evoked gamma band activity. This notion is aligned with the idea that first-order boundaries are encoded in a feedforward signal that exhibits high inter-trial phase coherence of gamma band activity. By contrast, second-order acoustic boundaries might not be encoded in the initial feed forward signal in low-level auditory pathways. Meaning that while first-order boundary information may be processed fully in the initial feedforward sweep to the auditory cortex, second-order boundaries may require integration of multiple areas in horizontal or feedback processing.

2.6 Experiment Two

Previous research on auditory boundaries identified a component of the ERP associated with sound onsets defined by the spectrotemporal discontinuity associated with changes in motion trajectory. An interesting aspect of the lateralized object-related negativity reported by Butcher et al. (2011) was that the LORN associated with motion-defined (i.e. second-order) objects appeared contralateral to the hemispace in which new auditory objects appeared. This was in contrast to first-order onsets, which exhibited midline N1 peaks and somewhat lateralized radial N1 peaks. It is possible that the N1

evoked by second-order boundaries shown in Experiment 1 is related to the LORN described by Butcher et al. If so, it should exhibit contralaterality like the LORN. Thus, we carried out a second experiment to determine if the N1 evoked by second-order IRN boundaries exhibits a contralateral focus relative to the side on which sound objects appear. A second goal of Experiment 2 was to replicate the early, evoked gamma signal observed in Experiment 1. Thus, Experiment 2 was similar to Experiment 1 except that only pitch-defined (IRN) targets were presented and these appeared at speakers to the left or right of the acoustic midline. Our theory for Experiment 2 was that even when these first-order and second-order boundaries are lateralized, the neural mechanisms required in processing them would be different. We predicted that second-order boundaries would exhibit a negative deflection on the contralateral side to the side of stimulus presentation, where first-order boundaries would not.

2.7 Experiment Two Methods

2.7.1 Participants

Fifteen students from the University of Lethbridge participated in Experiment 2 for course credit; none from Experiment 1. The nature of the experiment was explained to the participants. Written informed consent was completed and the participants reported no hearing or neurological problems. Procedures were in accordance with the Declaration of Helsinki and approved by the University of Lethbridge Human Subjects Review Committee. Trials contaminated with eye artifact (>120 μV) were rejected and data from subjects with less than 50% of original trials were not analyzed leaving twelve participants (three male; mean age: 25.3; one left-handed).

2.7.2 Stimuli

The first-order and second-order boundaries were created in the same manner as Experiment 1, except that the IRN stimuli had 8 iterations of a time delay of 40 samples, given a 32000Hz sample rate, resulting in a 1.25ms time delay. The resulting IRN stimuli had a percept corresponding to an 800Hz pitch. The increased number of iterations increased the salience of the IRN compared to the stimulus used in Experiment 1.

Experiment 2 had two stimulus conditions: silence-to-IRN and noise-to-IRN. The IRN-to-noise condition of Experiment 1 was not included. Stimuli were presented from speakers at 60° to either side of the participant's midline. The IRN in both conditions was only presented, pseudorandomly, at one of the two speakers while the other speaker played continuous broadband noise in the noise-to-IRN condition and was entirely silent for the silence-to-IRN condition. Blocks consisted of repeating cycles of silence alternating with IRN (first-order) or broadband noise alternating with IRN (second-order). Boundaries were jittered so that the duration of each stimulus was between 1250ms and 1750ms (Fig. 2.5). Thirty blocks of each condition were presented in pseudorandom order. Participants responded to IRN for both the silence-to-IRN condition and the noise-to-IRN condition by indicating which side the IRN appeared on (i.e. IRN onsets were the targets in a spatial discrimination task).

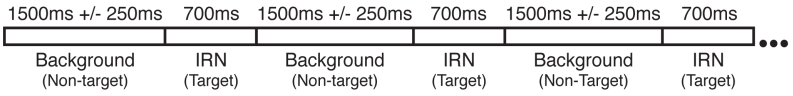


Figure 2.5: Stimulus presentation of Experiment 2. Ten cycles of 1250ms to 1750ms of broadband noise followed by 1250ms to 1750ms of IRN and ending with 1250ms to 1750ms of broadband noise.

2.7.3 Setup

The experimental setup was the same as that for Experiment 1.

2.7.4 Recording and Analysis

Recording followed the same procedure as Experiment 1, including sampling rate, impedances, and applied filters.

2.7.5 Behavioral Data

Average response times for Experiment 2 were computed for each condition in MATLAB. Only trials on which the participant correctly identified the IRN were included in these averages. The data were analyzed using a two boundary type (silence-to-IRN & noise-to-IRN) by two stimulus side (left & right) repeated-measures ANOVA.

2.7.6 Event-Related Potential Waveforms

For Experiment 2 we computed the ERP waveforms over a total epoch of 600ms, time-locked to spectrotemporal boundaries (onsets only). The waveforms were grand-averaged for each condition with an epoch of -200ms to 400ms using a baseline of 200ms to 0ms. Only trials in which the participant correctly identified the IRN were included in the average. Mean amplitudes (spanning +/- 8ms) of the N1 peak at lateral electrodes were analyzed in a two boundary type (silence-to-IRN & noise-to-IRN) by two stimulus side (left & right) by two electrode side (C5 & C6) repeated measures ANOVA. Scalp topography of the components in the grand-averaged waveforms was visualized with isopotential maps.

2.7.7 Time-Frequency Analysis

The time-frequency plots and permutation methods were identical to that of Experiment 1. The frequency bins were incremented by 2Hz between 4 and 60Hz as with Experiment 1. However, the time bins were incremented by 25ms from -200 to 400ms. We computed ITPC, total power, induced power, and evoked power at electrode Fz as in Experiment 1. To consider lateralization in the time-frequency data, we collapsed signals of single-trial data into contralateral (at C5 for right-side stimuli and at C6 for left-side stimuli) and ipsilateral (at C5 for left-side stimuli and at C6 for right-side stimuli). We computed ITPC, total power, induced power, and evoked power for contralateral and ipsilateral activity at electrodes C5 and C6 as in Experiment 1.

2.8 Experiment Two Results

2.8.1 Behavior

For Experiment 2, average response times for first-order boundaries were 410.89ms and 395.09ms for left and right side stimuli, respectively. Average response times for second-order boundaries were 414.86ms and 404.56ms for left and right side stimuli, respectively. There was no significant interaction between boundary type and stimulus side ($F_{1,11}$=2.100; p=0.175) for reaction time. There was also no significant effect due to boundary type ($F_{1,11}$=0.766; p=0.400) and only a marginal significant effect due to stimulus side ($F_{1,11}$=4.375; p=0.060) for reaction time. Average hit rate for first-order boundaries were 99.0% and 99.5% for left and right side stimuli, respectively. Average hit rate for second-order boundaries were 98.8% and 98.8% for left and right side stimuli, respectively. There was no significant interaction between boundary type

and stimulus side ($F_{1,11}$=1.574; p=0.236) for hit rate. There were also no significant effects of boundary type ($F_{1,11}$=1.385; p=0.264) or stimulus side ($F_{1,11}$=1.274; p=0.283) on average hit rate.

2.8.2 Event-Related Potentials & Isopotential Voltage Maps

ERP analysis for Experiment 2 showed auditory N1 deflections similar to those in Experiment 1. The N1 peak was maximal over fronto-central scalp and was slightly higher in amplitude for second-order boundaries compared to first-order boundaries, but was not significant ($F_{1,11}$=1.054; p=0.327). The N1 peak at electrode Fz occurred at 124ms for first-order boundaries, regardless of stimulus side, and at 144ms for second-order boundaries, regardless of stimulus side. There was an interaction between boundary type and stimulus side on N1 amplitude at electrode Fz ($F_{1,11}$=27.830; p<0.001). However, there was no main effect of stimulus side ($F_{1,11}$=2.090; p=0.176) on N1 amplitude. At electrode Fz, there was no boundary type by stimulus side interaction on N1 latency ($F_{1,11}$=0.278; p=0.608). As with Experiment 1, there was an effect of boundary type on N1 latency indicating that first-order N1 responses occurred earlier than the N1 to second-order boundaries ($F_{1,11}$=31.799; p<0.001). There was no effect of stimulus side on the N1 latency ($F_{1,11}$=1.121; p=0.312).

The early P90 deflection evoked by first-order, but not second-order boundaries, was also replicated in this experiment. ERP overplots for first- and second-order boundaries, collapsed across left and right side stimuli, are shown in Figure 2.6a.

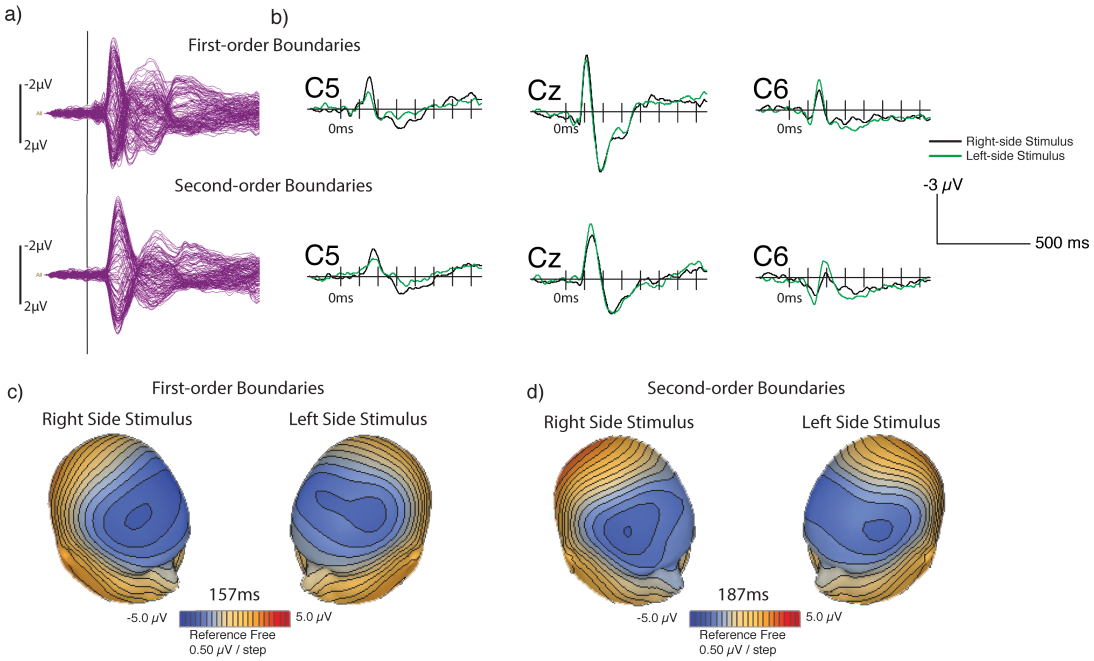


Figure 2.6: (a) ERP waveforms collapsed across left and right stimulus side, revealing the early P90 for first-order boundaries (top), but not second-order boundaries (bottom). (b) ERP waveforms at C5, Cz, and C6. C5 and C6 reveal increased amplitude at the electrode contralateral to the stimulus side for both first-order (top) and second-order (bottom) boundaries. (c) Lateralization is evident in the isopotential voltage maps evoked by first-order onsets at a latency of the N1 peak at electrodes C5 and C6 - 157ms. (d) Lateralization is evident in the isopotential voltage maps evoked by second-order onsets at the latency of the N1 peak at electrodes C5 and C6 - 187ms.

To visualize lateralization of ERP components, we plotted waveforms for electrodes C5 and C6. Butcher et al. (2011) reported that components evoked by second-order boundaries were relatively more lateralized than those evoked by first-order boundaries. Therefore, we expected to find a three-way (boundary type by stimulus side by electrode side) interaction. However, this interaction was not significant ($F_{1,11}$=2.094; p=0.176). Lateral electrodes, C5 and C6, show an N1 amplitude increase for stimuli presented from the speaker contralateral to the electrode position, regardless of boundary type (Fig. 2.6b). This is also revealed in the significant interaction of stimulus side by

electrode side interaction ($F_{1,11}$=23.951; p<0.001) and in the isopotential voltage maps. The N1 deflection at lateral electrodes C5 and C6 was later than the N1 deflection at electrode Fz. Therefore, the isopotential maps are shown for the N1 peak at these electrodes, 157ms for first-order boundaries (Fig. 2.6c) and 187ms for second-order boundaries (Fig. 2.6d). There were marginally significant interactions between boundary type and electrode side ($F_{1,11}$=4.198; p=0.065) and boundary type and stimulus side ($F_{1,11}$=3.430; p=0.091) on the mean N1 amplitude at these latencies.

2.8.3 Time-frequency Analysis

2.8.3.1 Electrode Fz

2.8.3.1.1 Theta and Alpha Bands

Consistent with Experiment 1, time-frequency analysis revealed ITPC and evoked power in the alpha and theta bands for both first- and second-order boundaries at electrode Fz (Fig. 2.7 rows 1-2, columns 1&4). The bottom row of Figure 2.7 shows the results of the permutation test comparing first-order and second-order responses. These data are masked to show only bins that met FDR correction threshold. There were no significant differences revealed in the induced power plots (Fig. 2.7, column 3).

2.8.3.1.2 Gamma Band

Again consistent with Experiment 1, we observed a transient evoked high-frequency signal maximal in the gamma band only for first-order boundaries (Fig. 2.7, row 1).

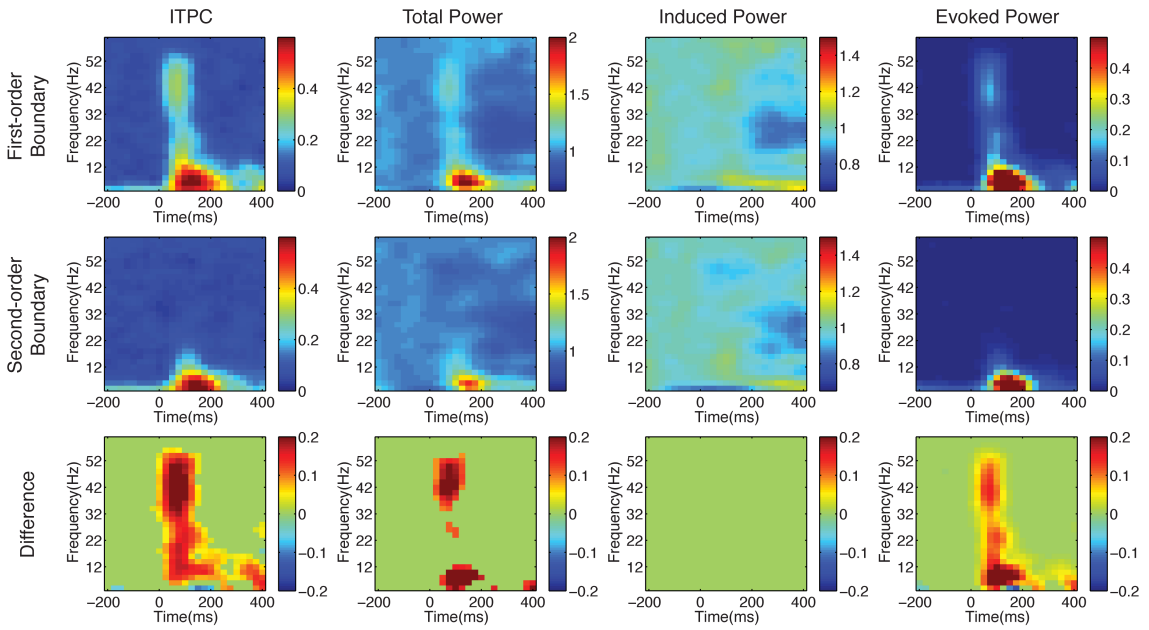


Figure 2.7: Time-frequency analysis of transitions at Fz when collapsed across left and right side stimulus. Inter-trial phase coherence (ITPC), total power, induced power, and evoked power are shown, respectively, in columns. Row 1: First-order boundaries; Row 2: Second-order boundaries; Row 3 displays the difference between the above conditions, masked to preserve false-discovery rate. The value of each bin is displayed only if that bin's p-value in the permutation test was less than the FDR cut-off. Bins that did not pass this FDR test are set to green.

2.8.3.2 Electrodes C5 and C6

2.8.3.2.1 Theta and Alpha Bands

In order to look further at lateralization, we performed time-frequency analysis at electrodes C5 and C6. These electrodes exhibited the largest lateralization of the N1 ERP component. We collapsed the data to contrast contralateral (at C5 for right-side stimuli and at C6 for left-side stimuli) with ipsilateral activity (at C5 for left-side stimuli and at C6 for right-side stimuli). First-order boundary responses are shown in row 1 of Figure 2.8 and second-order responses are shown in row 2. Evoked theta and alpha band activity to first-order boundaries was stronger on the contralateral side at the N1 latency compared to the ipsilateral side (Fig. 2.8; row 3), as revealed by ITPC, total, and evoked

power plots. The bottom row of Figure 2.8 shows the results of the permutation test comparing contralateral and ipsilateral activity. These data are masked to show only bins that met FDR correction threshold. There were no significant differences revealed in the induced power plots (Fig. 2.8, column 3). Data for second-order boundaries revealed similar activity at the theta and alpha bands (not shown).

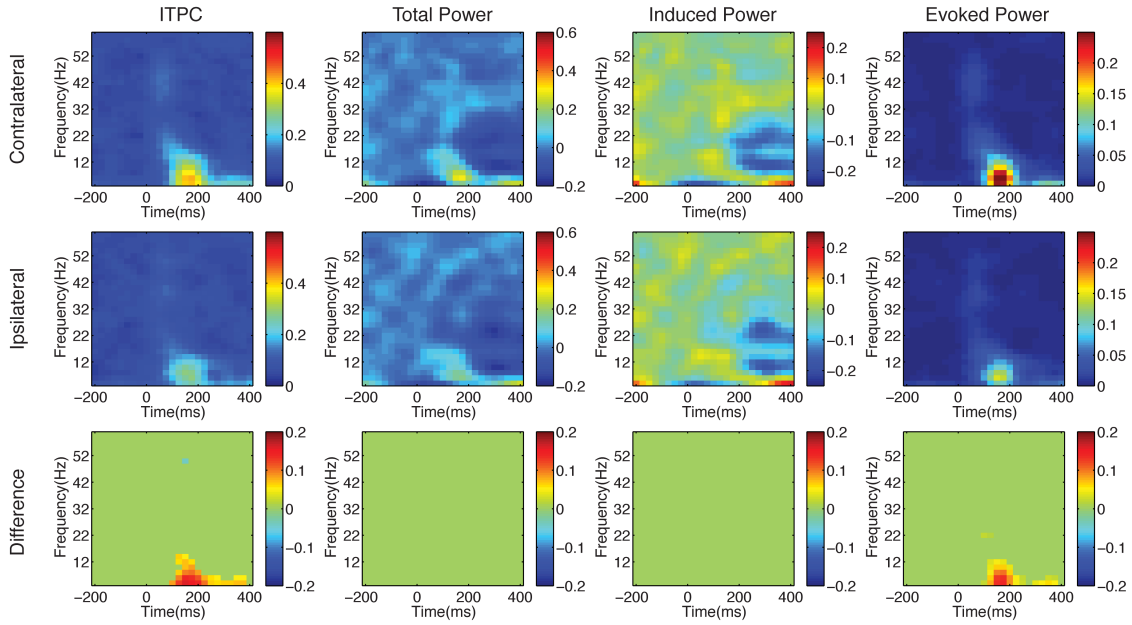


Figure 2.8: Time-frequency analysis of transitions at contralateral and ipsilateral electrodes. Inter-trial phase coherence (ITPC), total power, induced power, and evoked power are shown, respectively, in columns. Row 1: Contralateral signal; Row 2: Ipsilateral signal. Row 3 displays the difference between the above conditions, masked to preserve false-discovery rate. The value of each bin is displayed only if that bin's p-value in the permutation test was less than the FDR cut-off. Bins that did not pass this FDR test are set to green.

2.8.3.2.2 Gamma Band

The gamma activity is less evident in the contralateral and ipsilateral time-frequency plots because the gamma activity was maximal at mid-line electrodes. The evoked gamma activity was evident at 75ms in the evoked power plots only when the

data are masked to show p-values smaller than 0.05, but is absent with the more conservative FDR corrected values (Fig. 2.8, row 3).

2.9 Experiment Two Discussion

The goal of Experiment 2 was to investigate whether second-order IRN boundaries exhibit more lateralization than first-order boundaries with respect to ERP and time-frequency analysis. This goal followed from the report of lateralized object-related potential observed when auditory objects are defined by discontinuous motion. Butcher et al. (2011) panned different sounds around a listener in order to create motion-defined objects in the auditory scene. These sounds included a single sweep (left to right or vice versa), two sweeps in the same direction with no gap of silence between them (both left to right or right to left), and two sweeps in the opposite directions with no gap of silence between them (left to right followed by right to left). When the sound disappeared on one side and simultaneously reappeared on the other, they observed a negative deflection only on the side contralateral to the second sweep. They called this the lateralized object related negativity (LORN).

In the present study, ERP waveforms with prominent N1 potentials were elicited mainly at medial frontal electrodes. Lateralization of the N1 was observed at electrodes C5 and C6. Unlike the LORN reported by Butcher et al. (2011), ERP waveforms in the present study were of approximately the same amplitude for first-order and second-order boundaries. However, the present study did replicate the contralateral nature of these components. This lateralization is interesting in the case of second-order boundaries because the auditory scene in this condition was symmetric: the IRN stimulus in one

hemispace was balanced by broadband noise in the other hemispace. Thus, the total acoustic energy of the auditory scene was symmetric, while the IRN target appeared on only one side. The contralaterality observed here for second-order boundaries is not driven by an asymmetry in the acoustic scene, but is instead associated with processes subsequent to initial parsing of the auditory scene for object boundaries.

One interpretation of the differences between these data and the LORN reported by Butcher et al. (2011) is that the lateralized stimuli in Experiment 2 evoked a LORN regardless of boundary type. Butcher et al. suggested that the LORN might reflect the need to orient attention through space in response to events occurring off the auditory midline. In Experiment 2, the target side was unpredictable and thus probably triggered an attention-orienting response regardless of the nature of the boundary that defined it.

An alternative interpretation is that the lateralized IRN stimuli used in the present experiment did not trigger the same scene analysis mechanisms as the motion-defined stimuli used by Butcher et al. (2011). The motion-defined stimuli in their experiment jumped suddenly from one hemispace to the opposite hemispace. Therefore, an important characteristic of that auditory scene was a sudden discontinuity in the symmetry of total acoustic energy: an energy decrease in one hemispace was accompanied by a simultaneous energy increase in the other. In the present experiment, first-order boundaries had an energy increase in one hemispace, without a sudden decrease in the other hemispace. Likewise, second-order boundaries had a spectral change in one hemispace, while the total acoustic energy remained balanced across the midline. Finally, it is possible that only processes engaged when the brain is processing auditory spatial motion elicit the LORN reported by Butcher et al.

The secondary goal of Experiment 2 was to replicate the observation of a gamma signal with tight phase locking to first-order boundaries. As in Experiment 1, we found significant ITPC evoked by first-order, but not second-order boundaries. Experiment 2 further characterized this transient gamma signal by showing that it does not appear to lateralize with respect to the stimulus side in the same way as the later, lower-frequency components (i.e. the N1).

2.10 Conclusion

The mechanisms of the auditory system that parse the scene for objects and their boundaries have been associated with distinct components of the auditory evoked potential. The most prominent and well-studied of these are the object-related negativity (ORN) (Alain *et al.*, 2001b; Alain *et al.*, 2002) and the pitch onset response (POR) (Krumbholz *et al.*, 2003; Seither-Preisler *et al.*, 2004; Seither-Preisler *et al.*, 2006a; Seither-Preisler *et al.*, 2006b). The IRN stimulus itself is well suited to investigate object-related activity in the cortex because it triggers the appearance of a salient object in the auditory scene without an accompanying broad-band energy transient at the periphery. Since the IRN stimulus differs from the broadband background by virtue of its spectral properties, its most salient feature is that of pitch. In this sense the IRN stimulus is similar to the mistuned harmonic stimuli (Alain *et al.*, 2002) and dichotic pitch stimuli (Johnson *et al.*, 2003; Johnson *et al.*, 2006) used to elucidate the ORN. In each of these cases, the perceived object(s) can be segregated from the background, or from other objects in the scene, on the basis of the perception of pitch. The spectrotemporal signatures of most real-world auditory objects are characterized by abrupt changes in acoustic energy across

many frequencies - that is, they are usually first-order. By contrast, the IRN stimulus, mistuned harmonics, and dichotic pitch are auditory objects defined by second-order boundaries. It is for this reason that brain electrical activity elicited by their onsets (or offsets) can be interpreted as being related to auditory perceptual mechanisms beyond the stage of simple registration of energy transients. In other words, IRN, mistuned harmonics, and dichotic pitch probably reveal similar mechanisms of auditory object formation.

Encoding auditory object boundaries is a computationally complex task that is engaged nearly continuously by the auditory system. These experiments suggest that different kinds of auditory boundaries engage different neural mechanisms as early as 75ms after stimulus onset. In particular, transitions from a spectrally complex IRN object to broadband noise, or from silence to IRN, triggered a positive deflection in the ERP at 90ms, which was prior to the N1 peak. This deflection was notably absent for transitions from noise to IRN, suggesting that boundaries defined only by the abrupt emergence of a spectral pattern must be encoded by mechanisms that require different steps. One possibility, consistent with current theories of visual object perception, is that such boundaries cannot be encoded by feed-forward mechanisms and instead requires interaction between primary and secondary auditory cortex. In this sense, the second-order IRN boundary might be encoded in a manner more similar to texture defined boundaries (Zipser *et al.*, 1996) or illusory contours (Lee & Nguyen, 2001) in that there may be a requirement of neurons in secondary visual cortex to project information back to primary visual cortex in order to completely process that type of boundary.

First-order IRN onsets were accompanied by an early (~75ms) broadband transient response focused in the gamma range. Such transient broadband responses are not uncommon and may be due to a wide variety of causes. Some of these, such as microsaccades (small, fast movement of the eyes from the point of focus), are not even cortical in their origin. It is striking however that this transient gamma response was associated strictly with first-order onsets. No other boundary type triggered evoked gamma of any duration or latency even though other boundary types did exhibit similar deflections in the ERP waveform prior to the N1. Therefore the neural origins and functional significance of this evoked gamma signal remain unknown. One possibility is that it reflects early coherent and stimulus-locked activity across a network of neurons that encode boundaries during a feed-forward sweep of information through the lowest levels of the auditory pathway. It will be of interest to test whether gamma coherence at early latencies imparts any down-stream perceptual consequences such as increased salience or accuracy.

2.11 Acknowledgments

The authors would like to thank Karla Ponjavic-Conte, Jarrod Dowdall, and Sebastian Pavlovic. Research was funded by an NSERC Canada Discovery Grant to Matthew S. Tata.

Chapter 3: Neuroelectric Signals Associated with Boundaries of Motion-Defined Sounds

3.1 Preface

In the previous chapter, we investigated neuroelectric responses to first-order and second-order boundaries using an iterated rippled noise stimulus. This stimulus was chosen because while first-order boundaries had energy boundaries, the second-order transitions had highly salient pitch boundaries. Experiment 1 in the previous chapter had three transitions silence-to-IRN, noise-to-IRN, and IRN-to-noise. This experiment found that the silence-to-IRN and IRN-to-noise transitions elicited a P90-N1-P2 ERP complex, where the noise-to-IRN only elicited the later N1-P2 complex. We speculated that the P90 is therefore associated with spectral increments that were not present in the noise-to-IRN transition. We also found that first-order boundaries alone elicited an early, evoked broadband transient. These data suggested that these boundaries are engaging different neural mechanisms as early as 75ms after stimulus onset.

Experiment 2 sought to investigate the lateralization of these first-order and second-order boundaries. As with Experiment 1, the P90 and the early, evoked transient were only elicited by the first-order boundary and neither of these components appeared to be lateralized. We found that the lateralization was elicited for both first-order and second-order boundaries, rather than only for second-order boundaries, as predicted. Our basis for more lateralization in response to second-order boundaries was from motion-defined boundaries. It may not be surprising that the responses to the pitch-defined boundaries did not elicit the same responses as those previously reported for motion-

defined boundaries. It is for this reason that we wanted to investigate whether the lateralization previously reported in response to motion-defined boundaries would be evident when first-order and second-order motion boundaries were more carefully defined. In addition, we wanted to further characterize first-order and second-order auditory boundaries in an acoustic scene.

3.2 Introduction

Auditory Scene Analysis is a computationally complex task that includes parsing objects from their background and representing the spectrotemporal boundaries that distinguish them (Bregman, 1990). While progress has been made with regard to similar mechanisms in vision, mechanisms associated with auditory object boundaries remain unclear. A useful first approximation has been to extend concepts from vision science by analogy to the auditory domain. For example, it is known that different aspects of a visual scene are handled by specialized structures, including distinct pathways dedicated to object identification and motion processing (Marr, 1982; Zeki *et al.*, 1991; Goodale & Milner, 1992). Similarly, putative 'what' and 'where' pathways have also been identified in the auditory system (Rauschecker & Tian, 2000; Alain *et al.*, 2001a; Kubovy & Van Valkenburg, 2001; Tian *et al.*, 2001; Tata & Ward, 2005a). This basic similarity in functional segregation across systems suggests that object perception in vision and hearing might follow broadly analogous themes with respect to computational mechanisms.

In vision, the boundaries between objects and their background can be detected by registering discontinuities in luminance, but also isoluminant discontinuities in color

(Shapley, 1990; Shapley, 2002), texture (Mareschal & Baker, 1998), or motion (Anderson, 1997). The terms 'first order' and 'higher order' have been attached to luminance boundaries and isoluminant boundaries, respectively. Although auditory boundaries are spectrotemporal in nature (Griffiths & Warren, 2002; 2004), it is possible to extend visual analogies to the auditory domain. First-order boundaries are those defined by increments or decrements of sound energy relative to a background noise floor, whereas higher-order boundaries are defined by spectrotemporal discontinuities despite a globally constant sound energy level. Such higher-order auditory boundaries have previously been investigated in dichotic pitch (Ungan *et al.*, 1989; Johnson *et al.*, 2003; Johnson *et al.*, 2006), mistuned harmonics (Alain *et al.*, 2001a; Alain *et al.*, 2001b; Alain *et al.*, 2002), energy-matched pitch (Krumbholz *et al.*, 2003; Seither-Preisler *et al.*, 2004; Seither-Preisler *et al.*, 2006b), transitions from organized to disorganized stimuli (Chait *et al.*, 2007a; Chait *et al.*, 2007b; Chait *et al.*, 2008), and discontinuous motion (Butcher *et al.*, 2011). In each of these studies, researchers have identified neuroelectric correlates that are thought to be attributable to encoding the higher-order boundary itself, rather than non-specific responses to first-order boundaries.

Although the auditory periphery is fundamentally tonotopic rather than spatiotopic, spatial information nevertheless contributes powerful cues for auditory scene analysis (Bregman, 1990; Darwin, 1997; McDonald & Alain, 2005). Butcher et al. (2011) investigated the neuroelectric correlate of auditory object boundaries defined by discontinuous motion in a virtual auditory space. In that study, sounds could jump instantly from one trajectory to another, leaving no gap of silence in the overall acoustic scene. Perceptually, this sounded like the disappearance of one sound and the appearance

of a new sound at a different location. The reason for that type of perception is because in an everyday auditory scene, object do not typically disappear from one location and reappear at another. Therefore, when the sound onsets again, it is likely that it is perceived as a new object. The globally continuous sound had been parsed into distinct objects on the basis of interaural time and level discontinuities - creating auditory boundaries notionally similar to visual boundaries defined by binocular disparity. This boundary triggered a distinct pattern in the ERP signal that they called the Lateralized Object Related Negativity (LORN).

Although the LORN component reported by Butcher et al (2011) resembled the well-studied N1 ERP component evoked by typical first-order boundaries, it differed in its distribution on the scalp. Whereas first-order onsets of sound evoked a midline N1 maximal near the vertex, the LORN evoked by motion-defined boundaries was more frontal and was strongly lateralized over the hemisphere contralateral to the appearance of the new sound. Thus, first-order and second-order boundaries in that study evoked distinct morphology of the ERP waveform and might reflect processes that are unique to perception of different kinds of auditory boundaries. On the basis of these differences, Butcher et al. suggested that first-order boundaries might be encoded early in the auditory pathway, while higher-order boundaries might require later, possibly re-entrant, interaction between multiple cortical areas. In the present study, we sought to replicate the contralateral distribution of ERP components evoked by motion-defined boundaries. We further predicted that adding first-order cues to motion-defined boundaries would eliminate the LORN components by enabling earlier parsing of the auditory scene. More specifically, by providing first-order transients along with motion cues, we predicted that

the contralateral LORN pattern described by Butcher et al. would revert to the midline "onsets" pattern. Thus, we focused on components of the ERP waveforms at frontocentral electrodes and lateral electrodes.

The present study had two secondary goals: First, auditory scene analysis is believed to be mediated in part by pre-attentive mechanisms. Therefore, we considered whether object-related components of the ERP were modulated by task relevance by comparing passive and active listening to the same auditory stimuli. Second, the neuroelectric dynamics of signals associated with encoding of motion-defined auditory boundaries have not been investigated. We used a time-frequency decomposition technique to characterize both phase and power parameters of the EEG signal during processing of these boundaries.

3.3 Methods

3.3.1 Participants

Thirty-seven students from the University of Lethbridge participated for course credit. The nature of the experiment was explained to all participants and written informed consent was obtained. Procedures were in accordance with the Declaration of Helsinki and approved by the University of Lethbridge Human Subjects Review Committee. The participants reported no hearing or neurological problems. Trials contaminated with eye artifact (>120μV) were rejected and data from subjects with less than 50% of original trials were not analyzed, leaving fourteen participants (five male; mean age: 21.2; all right-handed) in the passive listening group and fourteen participants (5 male; mean age: 20.2; one left-handed) in the active listening group.

3.3.2 Stimuli and Setup

Auditory stimuli consisted of a pure tone of 400Hz mixed with broadband noise. The stimulus was generated using Praat software (Boersma & Weenink, 2005). Participants sat in a sound attenuated room in the centre of a ring of twelve speakers (Mackie HR-624mk2 studio monitors). Vector-Based Amplitude Panning (VBAP) - a multi-speaker analog of stereo panning - was used to render sounds into the space around the listener. Sounds appeared pseudorandomly starting on the participant's left or right and then panned toward the midline. The sound followed an arc through seven of the speakers, corresponding to the front 180° and terminating at 90° into the opposite hemispace. This panning was repeated such that the sound appeared to sweep six times through the space in front of the listener. Each sweep was 1000ms in duration. We refer to this 6-sweep sequence as one trial.

Two different kinds of trials were possible: A first-order boundary trial consisted of sweeps with an interstimulus interval (ISI) of silence between each sweep. The duration of this ISI was jittered between 250ms and 750ms (rectangularly distributed). A second-order boundary trial consisted of sweeps without intervening gaps of silence. Thus, the offset on one side and the onset on the other side were simultaneous. Because the first sweep of a trial was always a first-order onset (i.e. it always included an increment in sound level from background), it was not used in the average for first-order or second-order boundaries. Four trials were grouped pseudorandomly into blocks, two starting on the left and two starting on the right. Each block was separated by rest breaks. The experiment consisted of 20 first-order blocks and 20 second-order blocks, which were ordered pseudorandomly. Thus, each participant heard 200 sweeps for each of the four

conditions (left/right x first-order/second-order). A video monitor was placed in front of the participants, slightly below the front speaker, and participants were instructed to keep their gaze on a fixation dot in the middle of the screen. Responses were made via a computer keyboard.

The participants were split into two groups: a passive listening group and an active listening group. The passive listening group was asked to attend to a visual task that was presented on the monitor in front of them. The fixation dot was replaced with a letter and they were to respond if the presented letter was a vowel (target present on 33% of trials). All auditory stimuli were presented while the participants were engaged in the visual task.

The active listening group heard identical auditory stimuli, except that a seventh sweep was occasionally inserted into a trial. Listeners were asked to attend to the sweeps and indicate, at the end of the trial, whether six or seven sweeps occurred. The seventh sweep, when present, was not included for ERP analysis. There was no visual task presented to this group.

3.3.3 Recording and Analysis

The EEG was recorded at a 500Hz sampling rate using 128 silver/silver chloride electrodes in a plastic net (Electrical Geodesics, Inc., Eugene, OR, USA) and EGI's Netstation acquisition software. Impedances were maintained below 100kΩ. Further analysis of the EEG was done with BESA software (Megis Software, Gräfeling, Germany) and customized MATLAB (MATLAB 2010b, The Mathworks Inc., 2010, Natick, MA, USA) tools based on the EEGLab toolbox (Delorme & Makeig, 2004) and

the Fieldtrip toolbox (Oostenveld *et al.*, 2011). EEG recording was visually inspected for bad channels and no more than eight electrodes were replaced with interpolated signal. Digital filters were applied to the EEG data: high-pass (0.5Hz, 6dB/octave) forward-phase and low-pass (30Hz, 24dB/octave) zero-phase Butterworth filters. Trials contaminated with eye artifact (>120 μV) were rejected from the average. The averaged ERPs were re-references and interpolated by BESA to a standard 81-channel average-reference montage (Luu & Ferree, 2005) based on the 10-10 system.

3.3.4 Event-Related Potential Waveforms

We computed the ERP waveforms over 1000ms epochs, time-locked to the start of each sweep. ERP epochs are usually computed relative to a pre-stimulus baseline. However, the interval preceding each sweep differed for the two boundary types: first-order events necessarily have a silent pre-stimulus interval, whereas second-order events were preceded by the previous sweep. To avoid confounding different baselines in the ERP with boundary type, we used the entire 1000ms epoch as the baseline. This effectively normalized the ERP to have a mean of zero in each condition. We computed mean amplitude for the P90 and N1 components for each condition at electrode Fz, which was the electrode with the most prominent peaks. However, the P2 component was most prominent at electrode Cz and was computed at that electrode. These peaks were analyzed in a two x two (boundary type by stimulus side) analysis of variance (ANOVA). N1 mean amplitude was also computed for electrodes F3 and F4 to investigate lateralization. Lateralization was analyzed in a two x two x two (boundary type by stimulus side by electrode side) ANOVA. The amplitude of each component was

analyzed by computing the mean amplitude under a 20ms latency window determined from the grand average. The ANOVAs at Fz and at F3/F4 were computed for the passive listening and the active listening groups. We then analyzed these mean amplitudes across experiments in a between groups comparison. Scalp topography of the N1 in the grand average at lateral electrodes, F3 and F4, were visualized with isopotential voltage maps.

3.3.5 Time-Frequency Analysis

We investigated inter-trial phase coherence (ITPC) and time-spectral evolution (TSE) of power from the single-trial raw data. The signal was transformed to time-frequency space through complex demodulation as described by Hoechstetter et al. (2004).

(1) $$Z_{k,t,f} = A_{k,t,f} \cdot e^{i\theta_{k,t,f}}$$

(2) $$ITPC_{t,f} = \left| \frac{1}{N} \sum_{k}^{N} e^{i\theta_{k,t,f}} \right|$$

(3) $$B_f = \frac{1}{n_t} \sum_{t_{baseline}}^{t_0} \frac{1}{N} \sum_{k}^{N} \left| A_{k,t,f} \right|^2$$

(4) $$TP_{t,f} = \frac{\frac{1}{N} \sum_{k}^{N} \left| A_{k,t,f} \right|^2}{B_f}$$

where $A_{k,t,f}$ is the amplitude of the complex valued result ($Z_{k,t,f}$) of the complex demodulation for trial k, frequency f, and time *t*; $\theta_{k,t,f}$ is the phase; N is the number of trials; B_f is the mean power calculated using the entire epoch; n_t is the number of time bins; $TP_{t,f}$ is the total power (deviation from mean power).

ITPC measures how well the phases of signals match over successive trials. For both the passive and the active listening groups, frequency bins were incremented by 2Hz from 4Hz-60Hz and time bins were incremented by 25ms from 0ms-1000ms. Plots for ITPC were computed for electrode Fz. To compare across boundary type, we used non-parametric, Monte Carlo permutation tests. The permutations were repeated 48000 times for each participant. To control for multiple comparisons, we applied a false-discovery rate (FDR) correction method as described by Benjamini and Hochberg (1995). Significant effects of boundary type in the active listening group and the passive listening group were found using an FDR corrected alpha value corresponding to α=0.05. Significant effects of listening task (comparing across experiment groups) were also found using an FDR corrected alpha value corresponding to α=0.05.

3.4 Results

3.4.1 Event-Related Potentials & Isopotential Voltage Maps

3.4.1.1 Passive Listening

ERP waveforms for the passive listening group are shown in Figure 3.1a. When the participants were asked to attend to the visual task, both first-order and second-order boundaries triggered N1 and P2 deflections. However, only first order boundaries triggered a robust early P90 deflection. This is evidenced by a significant effect of boundary type on the amplitude of the P90 at electrode Fz ($F_{1,13}$=25.452; $p<0.001$). There was no boundary type by stimulus side interaction and no effect of stimulus side on the P90. There was no boundary type by stimulus side interaction and no main effects of boundary type or stimulus side on the amplitude of the N1 at electrode Fz. The amplitude

of the P2 at electrode Cz was larger for first-order boundaries. This is evidenced by a main effect of boundary type on the amplitude of the P2 at electrode Cz ($F_{1,13}$=12.892; p=0.003). However, there was no boundary type by stimulus side interaction or main effect of stimulus side on the amplitude of the P2.

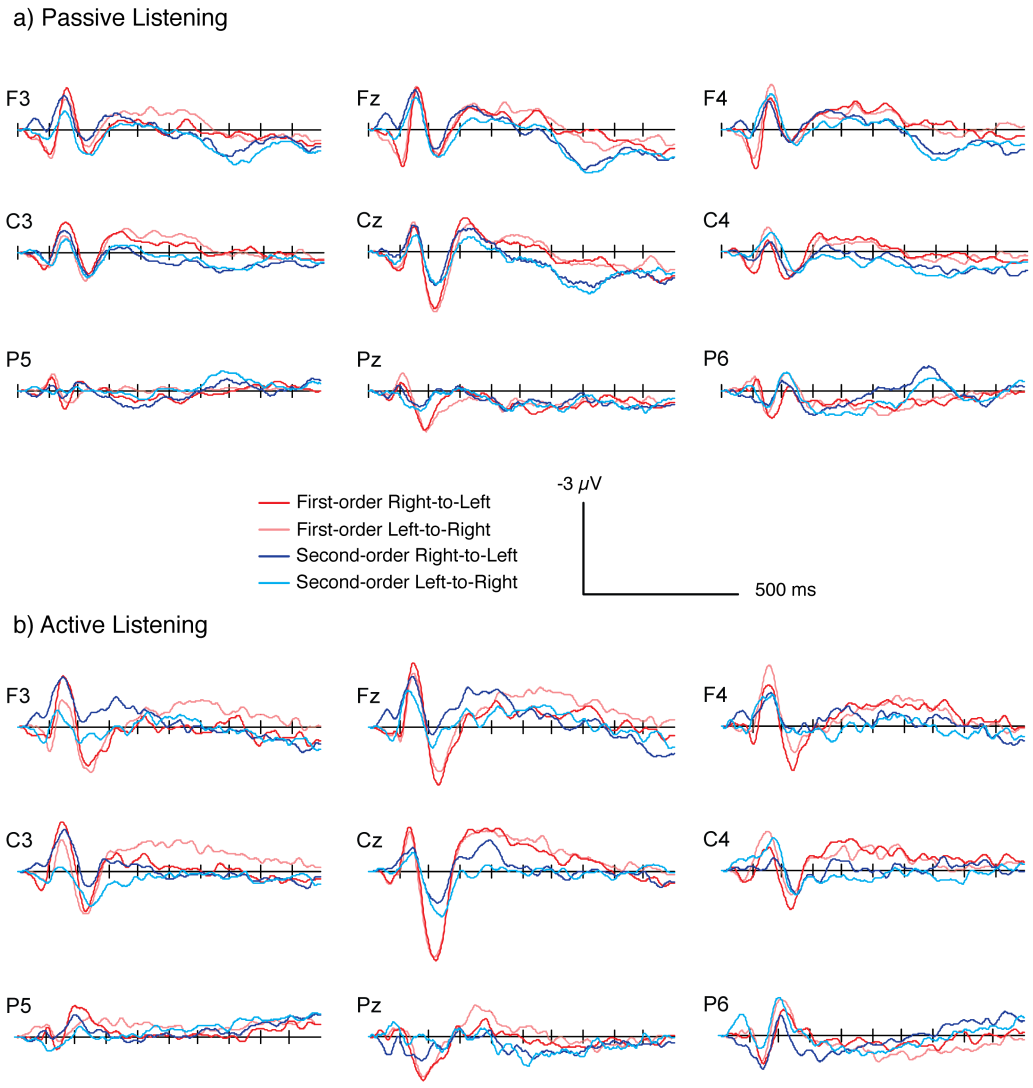


Figure 3.1: a) ERP waveforms to first-order right side start (red), first-order left side start (light red), second-order right side start (blue), and second-order left side start (light blue) for the passive listening group. b) ERP waveforms to first-order right side start (red), first-order left side start (light red), second-order right side start (blue), and second-order left side start (light blue) for the active listening group.

Figure 3.1a also shows the contralaterality of the evoked potentials at lateral electrodes, F3 and F4, for both first-order and second-order boundaries. The latency of the N1, calculated from the grand average, at these electrodes was 161ms for first-order boundaries and 157ms for second-order boundaries. There was a significant stimulus side by electrode side interaction ($F_{1,13}$=23.981; p=<001) indicating the contralaterality of this component - the predicted pattern that follows from Butcher et al. (2011). However the predicted three-way (boundary type x stimulus side x electrode side) interaction was not significant. That is, the pattern of contralaterality was similar, regardless of whether first-order cues were added to the motion boundary. Furthermore, no other two-way interactions of main effects were significant. Isopotential voltage maps at the latency of the N1 at F3 and F4 are shown in Figure 3.2a. These maps indicate foci over the frontal scalp on the contralateral side from where the stimulus appeared for both first-order and second-order boundaries.

Figure 3.2: a) Isopotential voltage maps shown for the N1 peak at electrodes F3 and F4 for the passive listening group. b) Isopotential voltage maps shown for the N1 peak at electrodes F3 and F4 for the active listening group.

3.4.1.2 Active Listening

ERP waveforms for the active listening group are shown in Figure 3.1b. As in the passive listening groups, both first-order and second-order boundaries triggered N1 and P2 deflections. However, neither first-order nor second-order boundaries triggered the early P90 deflection. There was no boundary type by stimulus side interaction and no effect of boundary type on the P90. The N1 deflection was greater in amplitude for first-order compared to second-order boundaries at electrode Fz. However, this effect was only marginally significant ($F_{1,13}$=3.291; p=0.093). There was no interaction of boundary type by stimulus side and no significant effect of stimulus side on the amplitude of the N1. As with the passive listening task, P2 deflections were most prominent at electrode Cz. At this electrode, first-order boundaries elicited a more robust P2 compared to second-order boundaries, which is evident in the significant main effect of boundary type ($F_{1,13}$=53.296; p<0.001). There was no boundary type by stimulus side interaction, nor was there a main effect of stimulus side on the amplitude of the P2.

Figure 3.1b also shows the contralaterality of both first-order and second-order boundaries in the active listening condition, which are revealed at lateral electrodes, F3 and F4. The latency of the N1, measure from the grand average at these electrodes, was 152ms for both first-order and second-order boundaries. There was a significant stimulus side by electrode side interaction on the N1 amplitude at these lateral electrodes ($F_{1,13}$=23.981; p<0.001), but no significant three-way interaction of boundary type by stimulus side by electrode side. Furthermore, there were no other main effects or interactions. As with the passive listening group, the interaction of stimulus side by electrode side indicates the contralateral nature of the ERP signals. However, the addition

of first-order cues did not cause the predicted change in lateralization pattern. Isopotential voltage maps at the latency of the N1 at F3 and F4 are shown in Figure 3.2b. As with the passive listening group, these maps indicate foci over the frontal scalp on the contralateral side from where the stimulus appeared for both first-order and second-order boundaries.

3.4.1.3 Across Experiment Groups

Across the passive and the active listening groups, there was a clear attentional effect for first-order boundaries, which was absent for second-order boundaries. At electrode Fz, there was a significant interaction of boundary type by listening task on the amplitude of the P90 ($F_{1,26}$=9.247; p=0.005) and a significant main effect of boundary type ($F_{1,26}$=23.018; p<0.001). These are both likely due to the P90 elicited for first-order boundaries in the passive listening task, but not to second-order boundaries in the passive listening group or either boundary type in the active listening group. There was an effect of stimulus side ($F_{1,26}$=11.705; p=0.002), which is likely due to the response to the second-order right-to-left stimulus eliciting a P90 in the opposite direction. There were no other significant interactions on the amplitude of the P90. There were no three-way interaction, no other interactions, and no effects on the amplitude of the N1. The amplitude of the P2 was larger for first-order boundaries in the active listening group compared to the second-order boundaries in the active listening group and both boundaries in the passive listening group. This was evidenced by a significant boundary type by listening task interaction on the amplitude of the P2 at electrode Cz ($F_{1,26}$=5.107; p=0.032) and a main effect of boundary type ($F_{1,26}$=57.381; p<0.001). There were no

other interactions or effects on the amplitude of the P2.

At lateral electrodes, F3 and F4, the effect of contralaterality was not significantly greater in the active listening compared to the passive listening task for either first-order or second-order boundaries. This is evidenced by the non-significant three-way stimulus side by electrode side by listening task comparisons for first-order ($F_{1,26}$=1.514; p=0.230) and second-order ($F_{1,26}$=0.802; p=0.379) boundaries.

3.4.2 Time-Frequency Analysis

3.4.2.1 Passive Listening

Time-frequency decomposition at electrode Fz for the passive listening group is shown in Figure 3.3a. Both boundary types triggered phase-locked power in the theta and alpha bands at the latency of the N1, which were greater due to first-order compared to second-order boundaries. This can be seen in the inter-trial phase coherence plots in column one. Results of the permutation test for the passive listening group are shown in row 3 of Figure 3.3a. These data are masked to show only the time-frequency bins with p-values less than the false-discovery rate (FDR). First-order boundaries triggered inter-trial phase coherence across multiple frequency bands approximately 50ms prior to the latency of the N1. When compared to the second-order boundary, this transient showed a significant ITPC difference mainly at the 42Hz (gamma) frequency range.

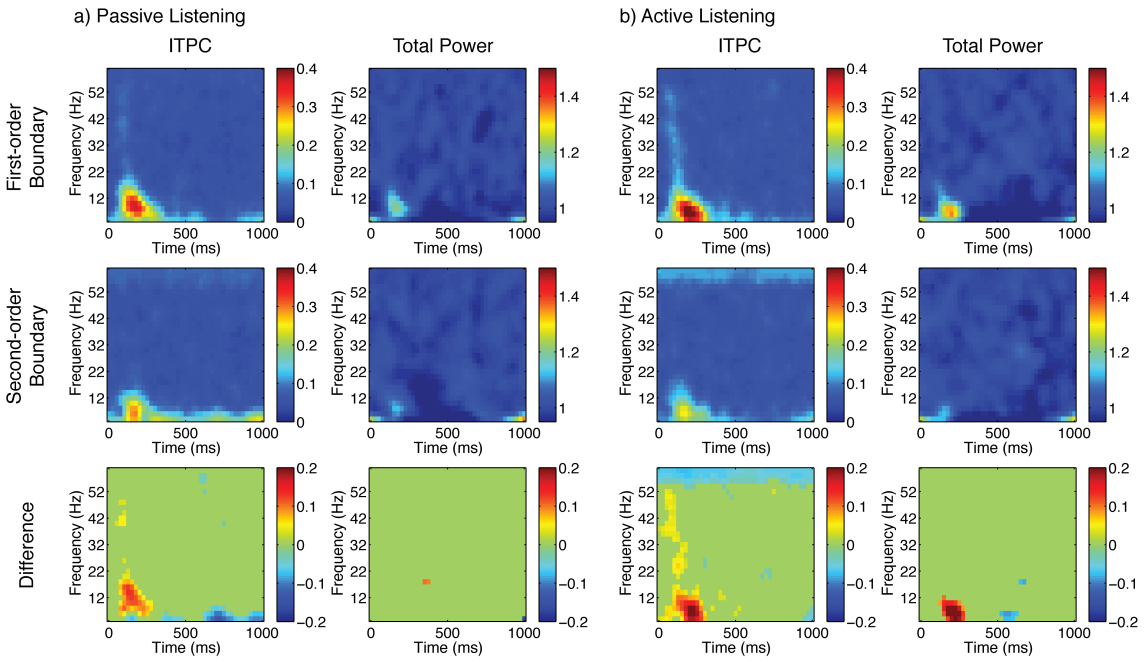


Figure 3.3: Time-frequency analysis for the passive (a) and active (b) listening groups. Inter-trial phase coherence in the left columns and total power in right columns. First-order boundaries are on the top rows and second-order boundaries are in the middle rows. The bottom rows displays the difference between boundary types, masked to preserve false-discovery rate. Bins that did not pass FDR are set to green.

3.4.2.2 Active Listening

Time-frequency decomposition at electrode Fz for the active listening group is shown in Figure 3.3b. Both boundary types triggered phase-locked power in the theta and alpha bands at the latency of the N1, which is shown by ITPC and evoked power. First-order boundaries also triggered early ITPC across multiple frequency bands. This ITPC was not triggered by second-order boundaries. The results of the permutation test comparing first-order and second-order boundaries in the active listening group are shown at the bottom of Figure 3.3b. These data are masked to show only the time-frequency bins with p-values less than the false-discovery rate (FDR). Unlike the passive listening group, the active listening group had significant ITPC across a wider range of frequencies - from ~25-50Hz.

3.4.2.3 Across Experiment Groups

Time-frequency analysis across the passive and active groups is shown in Figure 3.4a and 3.4b for first-order and second-order boundaries, respectively.

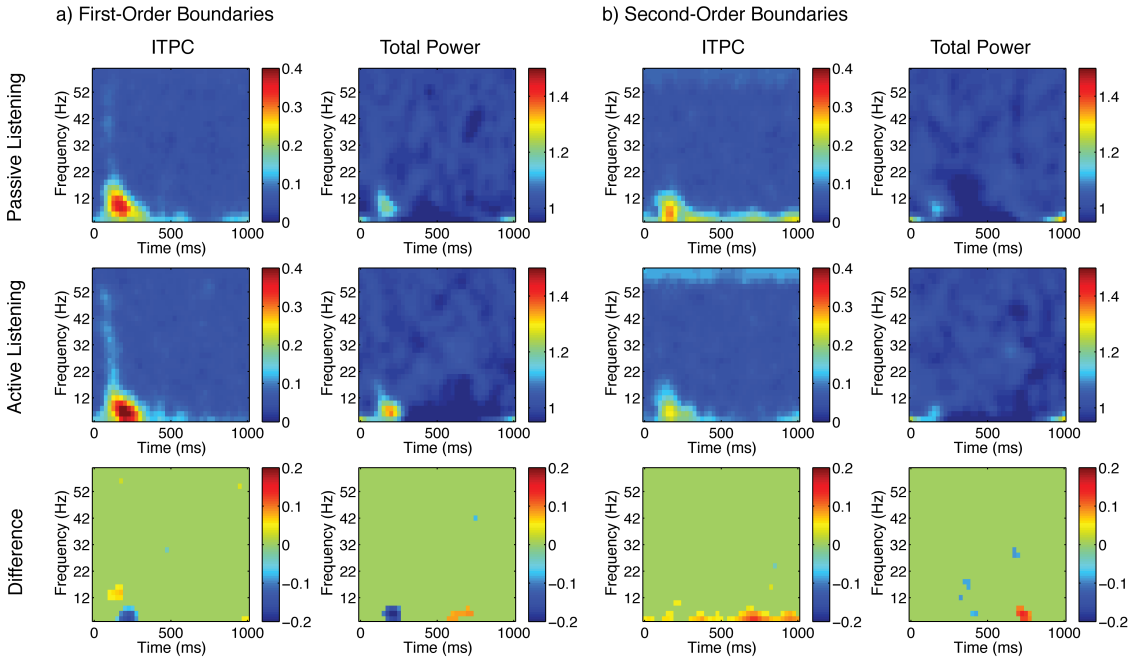


Figure 3.4: a) Time-frequency analysis for first-order (a) and second-order (b) boundaries. Inter-trial phase coherence in the left columns and total power in the right columns. Passive listening groups are on the top rows and active-listening groups are in the middle rows. The bottom rows displays the difference between listening tasks, masked to preserve false-discovery rate. Bins that did not pass FDR are set to green.

For first-order boundaries, there was ITPC in the passive listening group in the alpha/theta band earlier than ITPC in the active listening group. This may not be surprising considering there was an early P90 ERP deflection for the passive listening group and a later modulation of the P2 ERP deflection for the active listening group. However, when analyzing time-frequency plots, power can become blurred in time. Therefore, making conclusions about the latency of power or ITPC should be made lightly. Using a very conservative FDR mask, we observed no significant effect of task relevance on the early broadband transient centered around 40Hz. However, when less

conservative masking is used to show time-frequency bins with a corrected p-value of 0.05 or smaller, we do observe greater ITPC in the active listening group (not shown). The modulation was concentrated at the 30-35Hz range.

Figure 3.4b shows the time-frequency analysis in response to second-order boundaries. In the alpha/theta band, both first-order and second-order boundaries exhibited similar ITPC at the N1 latency. There was more phase-locked signal in the passive listening group in response to second-order boundaries that was spread out over the entire stimulus sweep.

3.5 Discussion

3.5.1 Event-Related Potentials

In the current study, we investigated auditory scene analysis in the context of neuroelectric responses to first-order and second-order auditory boundaries. A second-order auditory boundary that has received little consideration, but provides a strong cue for scene analysis, is a boundary defined by spatial dynamics - specifically, discontinuous motion. Such boundaries were thought to trigger a unique ERP correlate referred to as a lateralized object related negativity (LORN) (Butcher *et al.*, 2011). Our results suggest a somewhat different understanding of this ERP signal. We added first-order energy transients at the moment of a motion discontinuity and predicted that the presence of this extra cue would change the LORN pattern observed on the scalp. However, both first-order boundaries and second-order boundaries exhibited lateralized frontal negative peaks in the 100-200ms latency window. Thus, the frontal contralateral peak in the ERP is not unique to second-order boundaries. Instead, it may occur whenever a lateral cue to

object boundaries occurs, regardless of the type of cues that enable the brain to parse that boundary.

In the present study, we also considered whether the LORN might reflect an attentive or pre-attentive mechanism of scene analysis. We compared active and passive listening conditions and found that the LORN was relatively unaffected by task relevance. Thus, this ERP signal is probably a correlate of pre-attentive or reflexive mechanisms that are thought to parse the auditory scene prior to orienting of attention towards objects in that scene (Bregman, 1990).

In contrast to the LORN, we found two peaks in the ERP that did depend on the attentional set of listener. Both exhibited an interaction between task relevance and boundary type. A positive peak at about 90ms was greater for first-order boundaries, but only during the passive listening task. This P90 peak was maximal over frontal central scalp and did not exhibit substantial lateralization with respect to stimulus side. The P2 peak was greater for first-order boundaries only during the active listening task. This peak was maximal at Cz and also did not exhibit substantial lateralization relative to stimulus side.

The importance of task demands on the P90 and P2 might be because of differential effects of intermodal versus intramodal attention. Valtonen et al. (2003) investigated the effects of visual task load on the processing of unattended sounds using MEG. The magnetic analog of the N1 ERP component was found to increase in amplitude with increased visual task load. Similarly, Dyson et al. (2005) investigated how visual attentional load effects auditory ERP components. Like Valtonen et al., they found that there was modulation of the N1 ERP components. However, they categorized the

auditory ERP components into process-general (i.e. N1 and P2) and process-specific (i.e. the ORN) groups. While there was modulation of process-general components, visual task load failed to modulate more process-specific components. The first-order boundary results in the current experiment are in line with Dyson et al. in that there is modulation of the N1 and P2 due to visual task load. However, an important difference is that the current study compares visual task versus no visual task, where previous studies compare visual tasks that increase in difficulty. Therefore, differences in the ERP waveforms due to listening task in the current study may not exhibit the same modulation as those reported by Dyson et al.

Discontinuous motion stimuli investigated by Butcher et al. (2011) revealed that the LORN peak appeared approximately 65ms later than the N1 evoked by the initial first-order onsets of sound in the scene. One speculation by Butcher et al. was that first-order boundaries might be encoded in an early feed-forward sweep of information through the auditory system (Lamme & Roelfsema, 2000). By contrast, they suggested that second-order boundaries triggered the LORN due to feedback processes required for detecting such boundaries. In their design, second-order motion boundaries were compared to the onsets of sounds as they began their initial sweep through the acoustic space. In the present study, boundaries were not associated with onsets of sound, but rather discontinuities in the ongoing evolution of the acoustic scene. In this situation, the LORN peak has similar latency regardless of the boundary type. Therefore, it is unlikely that second-order boundaries in this case required more time than boundaries encoded by a first-order cue. Therefore, our data do not support Butcher et al.'s initial speculation regarding feed-forward vs. feedback processes, but rather suggest a difference between

the processing of onsets and the processing of boundaries subsequent to the original onset - regardless of the cues that comprise such boundaries.

Some ERP correlates of orienting attention into the auditory periphery have been investigated in previous studies. For example, Tata & Ward (Tata & Ward, 2005a) identified an early component of the mismatch negativity (MMN) triggered by unexpected deviations in the sound source. They used electrical source imaging to localize the generator of this MMN to posterior auditory cortex in the hemisphere contralateral to the deviant sound. Several other studies have reported a modulation of the ERP recorded over posterior electrodes due to spatially focused attention (Schröger, 1994; Schröger & Eimer, 1997; Tata *et al.*, 2001; Tata & Ward, 2005b). Tata et al. (2005b) used a combination of scalp topography and source imaging to show that an ERP generator in the hemisphere contralateral to the sound source was modulated by attention at about the same latency as the spatial MMN and the LORN. They suggested that this component might reflect orienting of attention on invalidly cued trials. More recently, Gamble & Luck (2011) reported a later component of the ERP that may reflect orienting attention into the periphery.

Butcher et al. (2011) speculated that the LORN might reflect a reorienting of attention similar to that triggered by a deviant (Tata & Ward, 2005a) or an invalid spatial cue (Tata & Ward, 2005b). Their reasoning was that a stable gestalt formed by continuous motion was broken and rebuilt due to the discontinuity of the stimulus. The LORN peak observed in the present study did not depend on boundary type. Thus, the LORN component appears not to require a voluntary orienting of attention. It could instead be a neural correlate of *reflexive* orienting of attention.

3.5.2 Time-Frequency Analysis

We conducted time-frequency analysis in order to investigate the inter-trial phase coherence and total power across a range of frequencies. In addition to phase-locked power in the alpha/theta band, the time-frequency results for first-order boundaries revealed a robust evoked transient across a large frequency range in both the passive and the active listening groups. Second-order boundaries also had phase-locked power in the alpha/theta band, but did not exhibit the same broadband evoked transient. Previous studies have described a transient gamma signal in auditory ERP waveforms that appeared when the ERP was filtered to show the 24-48Hz frequency range (Pantev *et al.*, 1991; Pantev & Elbert, 1994). Furthermore, evoked gamma-band signal has been modulated by stimulus intensity (Schadow *et al.*, 2007; McMullan *et al.*, under review), attention and task-relevance (Tiitinen *et al.*, 1993; Debener *et al.*, 2003), and anticipation (Schadow *et al.*, 2009). In these studies, the gamma band modulation was in response to auditory stimuli that are first-order boundaries. The current study showed that the broadband transient was evoked only for first-order boundaries and not for second-order boundaries. In addition, the transient was evoked for first-order boundaries in the active and the passive listening groups. The early, evoked broadband response to first-order boundaries may require an overall energy change or first-order boundary in the auditory scene. Meaning that it could reflect different neural signals that are elicited due to first-order aspects of an acoustic scene.

3.6 Conclusion

The dynamics of a time-varying auditory scene encode cues that enable the brain to parse the scene into discrete objects. We found that the presence or absence of a broad-spectrum energy transient (i.e. a first-order boundary) at the moment of a motion discontinuity modulates the response of the auditory system to that boundary. This modulation was evident both in components of the event-related potential and within specific frequency bands and latencies in the time-frequency decomposed EEG signal. First-order and higher-order boundaries exhibited different P90 and N2 peaks in the ERP. Furthermore, although first-order boundaries evoked larger lateralized frontal negative (LORN) peaks, the pattern of lateralization in the scalp topography was similar for both boundary types. Thus, this LORN ERP peak is not necessarily a unique correlate of second-order boundary processing. Furthermore, first-order boundaries triggered an early broadband evoked EEG signal and a stronger theta/alpha signal at the N1 latency relative to second-order boundaries. These results suggest that the brain does indeed encode first-order and second-order motion boundaries with distinct neural circuitry and mechanisms. The data also suggests that these mechanisms depend on the attentional set of the listener. However, the contralateral ERP correlates of boundary encoding are not strictly limited to higher-order boundary representations.

3.7 Acknowledgments

The authors would like to thank Dillon A. Hambrook for help with data collection and analysis. Research was funded by an NSERC Canada Discovery Grant to Matthew S. Tata.

Chapter 4: Discussion

4.1 General Discussion

The two studies in this thesis aimed to characterize the neuroelectric correlates of auditory scene analysis. More specifically, we sought to explore the mechanisms by which first-order and second-order auditory boundaries are processed. We predicted that the neural mechanisms required for this type of scene analysis would be different for first-order and second-order auditory boundaries. The research in the current thesis revealed the following interesting aspects of auditory scene analysis: First, different types of auditory boundaries engage different neural mechanisms. Second, first-order boundaries are encoded very early in the auditory pathway (by <90ms) whereas representations of higher-order boundaries may tale longer in the case of pitch, but not motion. Third, selective attention interacts with auditory boundary encoding at early stages of representation.

In Chapter 2, the auditory boundaries used were a first-order energy boundary and second-order boundaries with discontinuities in pitch. First-order boundaries were those characterized by an overall energy increment or decrement in the acoustic scene and second-order boundaries were those characterized by a discontinuity in the spectrotemporal envelope of the stimuli. Experiment 1 in Chapter 2 revealed differences in the ERP waveforms, including the P90 and the N1. Our theory for this experiment was that first-order boundaries would be processed earlier than second-order boundaries. As discussed in section 2.5, some predictions that could be made would be that the amplitude of the N1 would be greater for first-order boundaries compared to second-order

boudnaries and that listeners would be faster and more accurate in response to first-order boundaries. While the differences in the N1 amplitude in the ERP waveforms and the reaction times were not significantly different, the lack of significant shows that the responses to second-order boundaries were not simply a less robust first-order response. The early ERP component – the P90 – was present in response to transitions that had an increment in some or all frequency bands of the stimuli. These were the silence-to-IRN transition and the IRN-to-noise transition. It was the noise-to-IRN condition that did not have this early component. It could be that the change in the acoustic scene requires an increment at multiple frequency bands in order to elicit the P90.

An alternative interpretation of our data follows from a similarity between our stimuli and stimuli used by Chait and colleagues (Chait *et al.*, 2007a; Chait *et al.*, 2008). Chait et al. used a sequence of tone pips that were either constant pitch or varied randomly in pitch. When there was a transition from constant to random tones, deflections appeared in the MEG response at ~70ms and 150ms. Conversely, when there was a transition from random to constant tones, the MEG response showed only the second of these deflections (at ~150ms). Similar results were also found when comparing interaurally correlated stimuli to interaurally uncorrelated stimuli (Chait *et al.*, 2007b). Transitions from correlated to uncorrelated noise evoked deflections at ~80ms and 180ms, whereas transitions from uncorrelated to correlated noise evoked only a later response. Chait et al. have interpreted these data to mean that specific auditory mechanisms are triggered when the acoustic scene changes between ordered and disordered stimuli. Likewise, it is possible to interpret the data in the present study in the same context: a transition from noise to IRN could be interpreted as a transition from a

more disordered stimulus (noise) to a more ordered stimulus (IRN). However, it is hard to determine whether the brain considers silence, which has a small level of noise, to be more or less ordered than IRN. It could be that IRN is the more ordered stimulus compared to noise, but IRN is more disordered compared to silence. Alternatively, the critical factor may be the total area of the basilar membrane that is excited by a given stimulus over a brief period of time. A transition from a narrow-band stimulus to a broadband stimulus might trigger the early (P90) potential because it excited previously inactive regions of basilar membrane.

In addition, Experiment 1 in Chapter 2 revealed a broadband transient that was highly phase-locked and was only elicited in response to first-order boundaries. One possibility is that the transient gamma signal is an artifact of microsaccades triggered by abrupt acoustic transients (Yuval-Greenberg *et al.*, 2008). However, the other boundary type were no less salient, yet they did not trigger the transient evoked gamma signal. Another possibility is that the observed gamma transient is a broadband artifact of early components of the ERP, such as the P90, that happens to have relatively more power in the gamma band. However, the IRN-to-noise transitions also triggered a robust P90 component, but did not trigger transient evoked gamma. It could be that the ITPC reflects neural activity involved in representing events in the feedforward sweep of information, whereas second-order boundaries might not be encoded in this initial low-level auditory pathway. Thus, the transient high-frequency signal we observe is unique to first-order onsets and might reflect differential neural activity involved in representing such acoustic events.

Experiment 2 in Chapter 2 investigated the silence-to-IRN and noise-to-IRN transitions from Experiment 1, but lateralized the location from where the stimulus was presented. In regard to P90 and N1 differences in the ERP waveforms, this experiment revealed similar findings to those in Experiment 1. Namely, the P90 and the early, evoked gamma were elicited only in response to first-order boundaries. In addition, this experiment investigated a negative component – the LORN reported by Butcher et al. (2011) – that was present in response to second-order boundaries and not first-order boundaries. As with Butcher et al., this negativity in Experiment 2 of Chapter 2 was most robust at frontal electrodes contralateral to the side of the stimulus. However, the data in this experiment showed that the contralateral signal was elicited by both first-order and second-order boundaries. The lateralized stimuli were equally likely to appear to the left or right side of the listener and therefore, likely triggered attention orienting for both first-order and second-order boundaries. The P90 and evoked broadband ITPC elicited in response to only first-order boundaries suggest that first-order and second-order boundaries are likely engaging different neural mechanisms very early in processing. However, the LORN in the present study did not appear to depend on the whether the boundary was an energy increment or a change in the perceptual pitch of the stimulus.

Chapter 3 characterized the neuroelectric signals involved in encoding motion-defined auditory boundaries. First-order boundaries were those that had a silent gap at the time of the discontinuity and second-order boundaries did not. Furthermore, this study investigated correlates of these boundaries during an active and a passive listening task. The study in Chapter 3 revealed modulation of the P90 and P2 components due to task relevance. The P90 was greater in amplitude for first-order compared to second-order

boundaries during the passive listening task. The P2 was greater in amplitude in response to first-order compared to second-order boundaries during the active listening task. The P90 and P2 were both greater in response to first-order boundaries, suggesting that first-order and second-order discontinuous motion boundaries are encoded by different neural mechanisms. However, the modulation of these ERP components also depended on the listening task. The modulation of these components could also be due to the visual task load associated with the passive listening task compared to the active listening task, which was discussed in Chapter 3.

The study in Chapter 3 also sought to characterize an ERP component known as the LORN using more specific first-order and second-order discontinuous motion boundaries. Previously, the LORN due to motion-defined boundaries was elicited only by second-order boundaries (Butcher *et al.*, 2011). However, in Chapter 3, the LORN was elicited in response to both first-order and second-order discontinuous motion boundaries. This suggests a different interpretation from previous literature. Namely, that this component of the ERP may not require a second-order boundary. Instead, this component could reflect processing of initial onsets of an acoustic scene and boundaries that occur after an initial onset, regardless of what spectrotemporal properties comprise those subsequent boundaries. Alternatively, because this component is elicited when the acoustic stimuli jumps from one hemispace to the other, the LORN could reflect a reflexive orienting of attention.

This thesis sought to investigate the neuroelectric correlates of auditory scene analysis. First-order boundaries were defined as those that contained energy increments in both Chapter 2 and 3. Second-order boundaries were defined by discontinuities in pitch

in Chapter 2 and discontinuities in motion in Chapter 3. Both of these chapters indicated that first-order and second-order boundaries engage different neural mechanisms. One of the main indicators for this was that first-order boundaries elicited a broadband transient that was highly phase-locked to the boundary onset that occurred very early in processing – at about 75ms. In addition to this, first-order boundaries in Chapter 2 elicited an early P90, as did first-order boundaries in the passive listening task in Chapter 3. Again, indicating that these boundaries are engaging different neural mechanisms. Chapters 2 and 3 also investigated a frontal lateral component of the ERP, known as the LORN. The interpretation that the LORN reflects a reflexive orienting of attention could explain why this contralateral ERP component is elicited in response to first-order and second-order boundaries that are defined by discontinuities in motion and pitch.

4.2 Limitations of the Studies

The following two weaknesses can be identified in the approach we took to investigate auditory scene analysis: First, the definition of first-order and second-order auditory boundaries. Second, the electrodes used in the time-frequency analysis. For both of the chapters in this thesis, it was necessary to have a solid definition for auditory first-order and second-order boundaries. Primary auditory cortex (or the basilar membrane) is tonotopic. One key limitation is that there is the assumption that any change in frequency would cause a first-order boundary. It is the presence or absence of energy at a location on the basilar membrane that is the basic code for information in a sensory system. Processing leads to integration over time and space (across the retina or skin) or across time and frequency (across the basilar membrane). The analogy from vision may not be

the ideal comparison, but it gives a starting point to investigate and discuss auditory first-order and second-order boundaries.

In the studies in this thesis, we mostly reported time-frequency data at only one electrode. We showed ERP waveforms from multiple electrode sites and did investigate the time-frequency lateralization of the phase-locked gamma transient at lateral electrodes. The bulk of the time-frequency analysis was done at electrode Fz. This electrode was chosen for analysis based on the electrodes where the ERP components were most prominent. There is an apparent limitation in reporting the inter-trial phase coherence and power at only one electrode. In the initial stages of time-frequency analysis, we also focused on other electrodes, such as Cz, and found that the aspects of ITPC and power seen in the plots at Fz were also evident in the plots for Cz. However, these components were less robust at other electrodes. The goal of the thesis was to investigate the processing of first-order and second-order auditory boundaries. The predictions that followed from this were that first-order boundaries would be processed earlier than second-order boundaries. From this, the important aspect of the neuroelectric responses to these boundaries is the temporal properties. Meaning that, while placing emphasis on which electrode the signal comes from is important for characterizing correlates of these boundaries, it is secondary to the goal of the thesis.

4.3 Future Directions

This thesis consisted of investigating the neuroelectric correlates of auditory scene analysis. The boundaries used in the studies above included first-order energy boundaries, second-order pitch boundaries, and second-order discontinuous motion

boundaries. While there seemed to be a few ERP and time-frequency components that were consistent across the different experiments, several questions came to light. First, whether the early, broadband transient in inter-trial phase coherence elicited by first-order boundaries required the change in energy to reach a specific threshold. It could be the case that there is an all-or-nothing ITPC response, but it may also be the case that the energy increment from silence to a sound causes an associated ITPC response. Meaning that, as the energy increment gets larger, so does the ITPC response to that boundary. It would be of interest to further investigate the characteristics of the first-order boundary that cause the evoked broadband transient. In addition, whether the early, broadband transient has some perceptual consequences, such as increased reaction time or higher detection rate. As described in the discussion of Chapter 2, gamma signals have been associated with bottom-up and top-down processing (Herrmann *et al.*, 2010). In the current thesis, the tasks were simple detection task and listeners performed at ceiling levels for all boundary types. In future studies, the behavioral consequences of this gamma signal should be carefully considered.

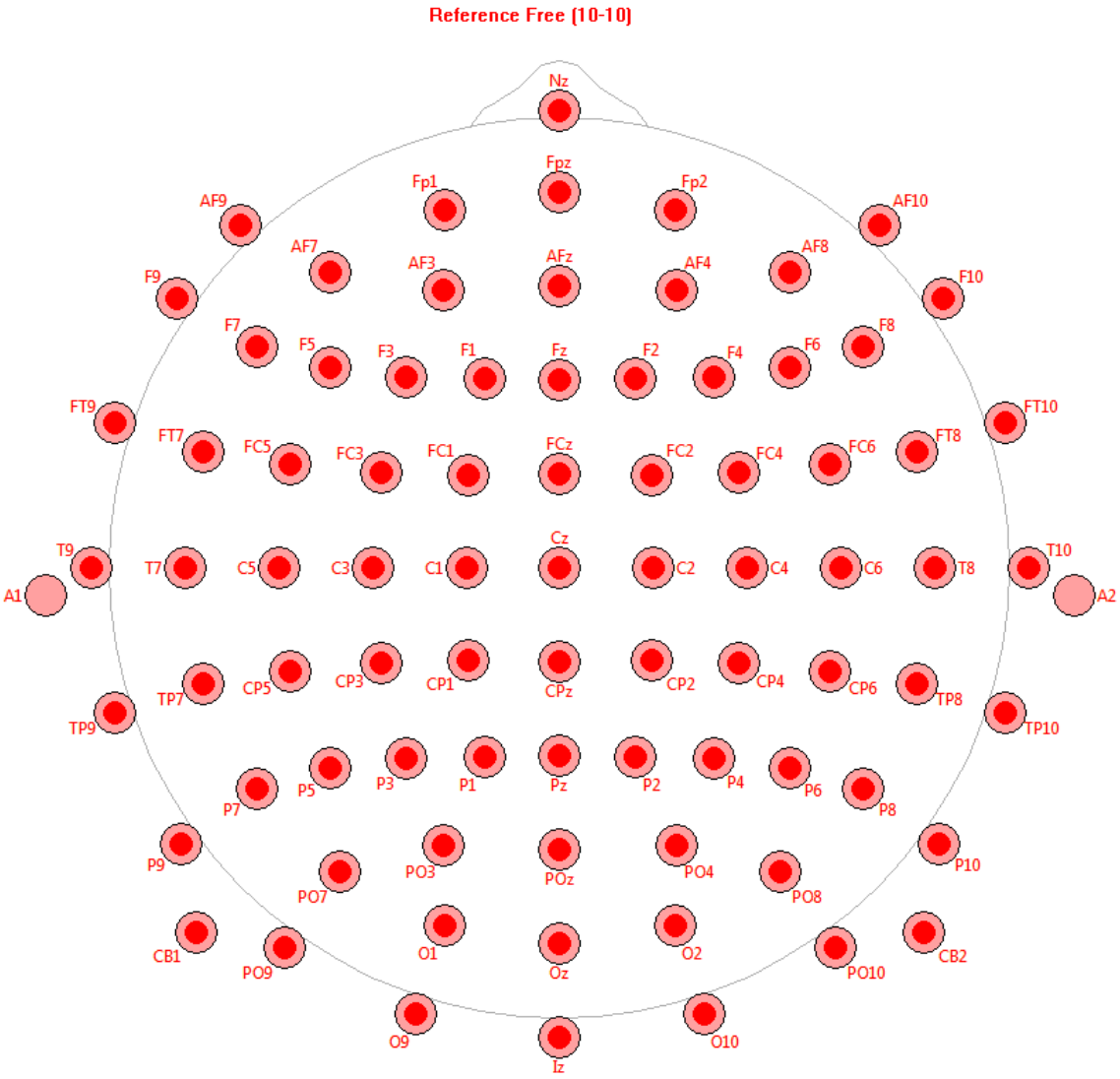
Secondly, whether neuroelectric responses are due to the perceptual aspects of the stimulus or whether they depend on the area taken on the basilar membrane. In Chapter 2, the responses to first-order energy boundaries and the second-order IRN-to-noise boundaries both elicited the early P90, where the noise-to-IRN boundary did not. One reason for this could be that this response is due to there being an activation of more area on the basilar membrane. It would be of substantial interest to investigate if that early P90 is only elicited when a stimulus excites a larger total area of the basilar membrane. If there is early processing based on the total area of the basilar membrane that is excited, it

could provide a fundamental characteristic of auditory processing. By investigating the correlates of simpler auditory scene, such as the ones in this thesis, it may be possible to eventually understand the ability of the brain to successfully parse more complex auditory scenes.

With the knowledge gained in the current thesis, a good launching point for more global investigation may have been established. These investigations may include whether the mechanisms of scene analysis are optimally engaged in more complex environments as well as the implications of disorders such as age-related hearing loss. The current thesis investigated the neuroelectric correlates of very simple auditory scenes. However, a person's every day auditory scene is much more complex in nature. Not only do typical auditory scene entail more complex sounds, such as music, they also entail complex speech. The neuroelectric correlates of these more complicated auditory scenes may provide an understanding of the auditory system that is currently lacking. With this understanding, it may be possible to further investigate age-related hearing loss. People with age-related hearing loss tend to be able to perceive and understand their auditory scene if they are in a relatively quiet environment. However, when they are in a more complex, or noisy, environment, their perception and understanding start to falter. It could be the case that people with this disorder are able to optimally encode first-order boundaries that are in a quiet environment, but are unable to encode the higher-order boundaries that are likely required in more complex auditory scenes. With a more complete understanding of how the brain encodes first-order and higher-order boundaries, it is conceivable that this would have an implication for better hearing aids and cochlear implants. The understanding of the auditory pathway and how auditory

boundaries are encoded is currently at the beginning stages, but as the correlates of these types of boundaries continue to be investigated, the implications appear to be important and could impact many people.

Appendix A: International 10-10 Electrode Placement Locations



References

Alain, C. & Arnott, S.R. (2000) Selectively attending to auditory objects. *Frontiers in Bioscience*, **5**, D202-212.

Alain, C., Arnott, S.R., Hevenor, S., Graham, S. & Grady, C.L. (2001a) "What" and "where" in the human auditory system. *Proceedings of the National Academy of Sciences*, **98**, 12301-12306.

Alain, C., Arnott, S.R. & Picton, T.W. (2001b) Bottom-up and top-down influences on auditory scene analysis: Evidence from event-related brain potentials. *Journal of Experimental Psychology*, **27**, 1072-1089.

Alain, C., Schuler, B.M. & McDonald, K.L. (2002) Neural activity associated with distinguishing concurrent auditory objects. *The Journal of the Acoustical Society of America*, **111**, 990-995.

Albright, T.D. (1992) Form-cure invariant motion processing in primate visual cortex. *Science*, **255**, 1141-1143.

Anderson, R.A. (1997) Neural mechanisms of visual motion perception in primates. *Neuron*, **18**, 865-872.

Baramidze, V., Lai, M.J. & Shum, C.K. (2006) Spherical splines for data interpolation and fitting. *SIAM Journal on Scientific Computing*, **28**, 241-259.

Benjamini, Y. & Hochberg, Y. (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B Methodological*, **57**, 289-300.

Boersma, P. & Weenink, D. (2005) Praat: Doing phonetics by computer.

Bosman, Conrado A., Schoffelen, J.-M., Brunet, N., Oostenveld, R., Bastos, Andre M., Womelsdorf, T., Rubehn, B., Stieglitz, T., De Weerd, P. & Fries, P. (2012) Attentional stimulus selection through selective synchronization between monkey visual areas. *Neuron*, **75**, 875-888.

Bregman, A.S. (1990) *Auditory Scene Analysis: The Preceptual Organization of Sound.* MIT Press, Cambridge, MA.

Busch, N.A., Schadow, J., Fründ, I. & Herrmann, C.S. (2006) Time-frequency analysis of target detection reveals an early interface between bottom-up and top-down processes in the gamma-band. *NeuroImage*, **29**, 1106-1116.

Butcher, A., Govenlock, S.W. & Tata, M.S. (2011) A lateralized auditory evoked potential elicited when auditory objects are defined by spatial motion. *Hearing Research*, **272**, 58-68.

Chait, M. (2005) Human auditory cortical processing of changes in interaural correlation. *Journal of Neuroscience*, **25**, 8518-8527.

Chait, M., Poeppel, D., de Cheveigne, A. & Simon, J.Z. (2007a) Processing asymmetry of transitions between order and disorder in human auditory cortex. *Journal of Neuroscience*, **27**, 5207-5214.

Chait, M., Poeppel, D. & Simon, J.Z. (2006) Neural response correlates of detection of monaurally and binaurally created pitches in humans. *Cerebral Cortex*, **16**, 835-848.

Chait, M., Poeppel, D. & Simon, J.Z. (2007b) Stimulus context affects auditory cortical responses to changes in interaural correlation. *Journal of Neurophysiology*, **98**, 224-231.

Chait, M., Poeppel, D. & Simon, J.Z. (2008) Auditory temporal edge detection in human auditory cortex. *Brain Research*, **1213**, 78-90.
Darwin, C.J. (1997) Auditory grouping. *Trends in Cognitive Sciences*, **1**, 327-333.
David, O., Kilner, J.M. & Friston, K.J. (2006) Mechanisms of evoked and induced responses in MEG/EEG. *NeuroImage*, **31**, 1580-1591.
Debener, S., Herrmann, C.S., Kranczioch, C., Gembris, D. & Engel, A.K. (2003) Top-down attentional processing enhances auditory evoked gamma band activity. *NeuroReport*, **14**, 683-686.
Delorme, A. & Makeig, S. (2004) EEGLAB: MATLAB toolbox for electrphysiological data analysis. Institute for Neural Computation, University of California at San Diego, San Diego (California).
Doesburg, S.M., Kitajo, K. & Ward, L.M. (2005) Increased gamma-band synchrony precedes switching of conscious perceptual objects in binocular rivalry. *NeuroReport*, **16**, 1139-1142.
Dyson, B.J. & Alain, C. (2004) Representation of concurrent acoustic objects in primary auditory cortex. *Journal of the Acoustical Society of America*, **115**, 280-288.
Dyson, B.J., Alain, C. & He, Y. (2005) Effects of visual attentional load on low-level auditory scene analysis. *Cognitive, Affective, & Behavioral Neuroscience*, **5**, 319-338.
Fries, P. (2005) A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends in Cognitive Sciences*, **9**, 474-480.
Fries, P., Schröder, J.-H., Roelfsema, P.R., Singer, W. & Engel, A.K. (2002) Oscillatory neuronal synchronization in primary visual cortex as a correlate of stimulus selection. *Journal of Neuroscience*, **22**, 3739-3754.
Gamble, M.L. & Luck, S.J. (2011) N2ac: An ERP component associated with the focusing of attention within an auditory scene. *Psychophysiology*, **48**, 1057-1068.
Glasberg, B.R. & Moore, B.C.J. (2002) A model of loundess applicable to time-varying sounds. *Journal of Audio Engineering Society*, **50**, 331-342.
Goodale, M.A. & Milner, A.D. (1992) Separate visual pathways for perception and action. *Trends in Neuroscience*, **15**, 20-25.
Griffiths, T.D. & Warren, J.D. (2002) The planum temporale as a computational hub. *Trends in Neuroscience*, **25**, 348-353.
Griffiths, T.D. & Warren, J.D. (2004) What is an auditory object. *Nature Reviews Neuroscience*, **5**, 887-892.
Hari, R., Pelizzone, M., Mäkelä, J.P., Hällström, J., Leinonen, L. & Lounasmaa, O.V. (1987) Neuromagnetic responses of the human auditory cortex to on- and offsets of noise bursts. *Audiology*, **26**, 31-43.
Herrmann, C.S., Fründ, I. & Lenz, D. (2010) Human gamma-band activity: A review on cognitive and behavioral correlates and network models. *Neuroscience & Biobehavioral Reviews*, **34**, 981-992.
Hillyard, S.A., Hink, R.F., Schwent, V.L. & Picton, T.W. (1973) Electrical signs of selective attention in the human brain. *Science*, **182**, 177-180.
Hillyard, S.A. & Picton, T.W. (1978) ON and OFF componenets in the auditory evoked potential. *Perception & Psychophysics*, **24**, 391-398.

Hoechstetter, K., Bornfleth, H., Weckesser, D., Ille, N., Berg, P. & Scherg, M. (2004) BESA source coherence: A new method to study cortical oscillations. *Brain Topography*, **16**, 233-238.

Hubel, D.H. & Wiesel, T.N. (1962) Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology*, **160**, 106-154.

Johnson, B.W., Hautus, M. & Clapp, W.C. (2003) Neural activity associated with binaural processes for the perceptual segregation of pitch. *Clinical Neurophysiology*, **114**, 2245-2250.

Johnson, B.W., Hautus, M.J., Hayns, A.L. & Fitzgibbon, B.M. (2006) Differential cortical processing of location and pitch changes in dichotic pitch. *NeuroReport*, **17**, 389-393.

Krumbholz, K., Patterson, R.D., Seither-Preisler, A., Lammertmann, C. & Lütkenhöner, B. (2003) Neuromagnetic evidence for a pitch processing center in Heschl's gyrus. *Cerebral Cortex*, **13**, 765-772.

Kubovy, M. & Van Valkenburg, D. (2001) Auditory and visual objects. *Cognition*, **80**, 97-126.

Kumar, S., Stephan, K.E., Warren, J.D., Friston, K.J. & Griffiths, T.D. (2007) Heirarchical processing of auditory objects in humans. *PloS Computational Biology*, **3**, 977-985.

Lamme, V.A.F. & Roelfsema, P.R. (2000) The distinct modes of vision offered by feedforward and recurrent processing. *Trends Neuroscience*, **23**, 571-579.

Lamme, V.A.F., van Dijk, B.W. & Spekreijse, H. (1994) Organization of contour from motion processing in primate visual cortex. *Vision Research*, **34**, 721-735.

Lee, T.S. & Nguyen, M. (2001) Dynamics of subjective contour formation in the early visual cortex. *Proceedings of the National Academy of Sciences*, **98**, 1907-1911.

Luck, S.J. (2005) *An Introduction to the event-related potential technique*. MIT Press, Cambridge, MA.

Lütkenhöner, B., Seither-Preisler, A., Krumbholz, K. & Patterson, R.D. (2011) Auditory cortex tracks the temporal regularity of sustained noisy sounds. *Hearing Research*, **272**, 85-94.

Luu, P. & Ferree, T. (2005) Determination of the HydroCel Geodesic Sensory Nets' average electrode positions and their 10-10 international equivalents. *Electrical Geodesics, Inc. Technical Note*.

Mareschal, I. & Baker, C.L.J. (1998) A cortical locus for the processing of contrast-defined contours. *Nature Neuroscience*, **1**, 150-154.

Marr, D. (1982) *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. W.H. Freeman, San Francisco.

McDonald, K.L. & Alain, C. (2005) Contribution of harmonicity and location to auditory object formation in free field: Evidence from event-related brain potentials. *The Journal of the Acoustical Society of America*, **118**, 1593.

McMullan, A.R., Hambrook, D.A. & Tata, M.S. (2013) Brain dynamics encode the spectrotemporal boundaries of auditory objects. *Under Review*.

Näätänen, R. & Picton, T. (1987) The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. *Psychophysiology*, **24**, 375-425.
Oostenveld, R., Fries, P., Meris, E. & Schoffelen, J.-M. (2011) FieldTrip: Open source software for advances analysis of MEG, EEG, and invasive electrophysiological data. In Neuroscience, C.I.a. (ed).
Pantev, C. & Elbert, T. (1994) The transient auditory evoked gamma-band field. In Pantev, C., Elbert, T., Lütkenhöner, B. (eds) *Oscillatory Event-Related Brain Dynamics*. Plenum, New York, pp. 219-242.
Pantev, C., Makeig, S., Hoke, M., Galambos, R., Hampson, S. & Gallen, C. (1991) Human auditory evoked gamma-band magnetic fields. *Proceedings of the National Academy of Sciences*, **88**, 8996-9000.
Pfefferbaum, A., Buchsbaum, M. & Gips, J. (1971) Enhancement of the average evoked reponse to tone onset and cessation. *Psychophysiology*, **8**, 332-339.
Polich, J. & Kok, A. (1995) Cognitive and biological determinants of P300: An integrative review. *Biological Psychology*, **41**, 103-146.
Ponjavic-Conte, K.D., Hambrook, D.A., Pavlovic, S. & Tata, M.S. (2013) Dynamics of distraction: Competition among auditory streams modulates gain and disrupts inter-trial phase coherence in the human electroencephalogram. *PLoS ONE*, **8**, e53953.
Rauschecker, J.P. & Tian, B. (2000) Mechanisms and streams for processing of "what" and "where" in auditory cortex. *Proceedings of the National Academy of Sciences*, **97**, 11800-11806.
Sauseng, P. & Klimesch, W. (2008) What does phase information of oscillatory brain activity tell us about cognitive processes? *Neuroscience & Biobehavioral Reviews*, **32**, 1001-1013.
Schadow, J., Lenz, D., Dettler, N., Fründ, I. & Herrmann, C.S. (2009) Early gamma-band responses reflect anticipatory top-down modulation in the auditory cortex. *NeuroImage*, **47**, 651-658.
Schadow, J., Lenz, D., Thaerig, S., Busch, N., Frund, I. & Herrmann, C. (2007) Stimulus intensity affects early sensory processing: Sound intensity modulates auditory evoked gamma-band activity in human EEG. *International Journal of Psychophysiology*, **65**, 152-161.
Schröger, E. (1994) Human brain potential signs of selection by location and frequency in an auditory transient attention situation. *Neuroscience Letters*, **173**, 163-166.
Schröger, E. & Eimer, M. (1997) Endogenous covert spatial orienting in audition: "Cost-benefit" analyses of reaction times and event-related potentials. *The Quarterly Journal of Experimental Psychology*, **50A**, 457-474.
Seither-Preisler, A., Krumbholz, K., Patterson, R., Seither, S. & Lütkenhöner, B. (2004) Interaction between the neuromagnetic responses to sound energy onset and pitch onset suggests common generator. *European Journal of Neuroscience*, **19**, 3073-3080.
Seither-Preisler, A., Patterson, R., Krumbholz, K., Seither, S. & Lütkenhöner, B. (2006a) Evidence of pitch processing in the N100m component of the auditory evoked field. *Hearing Research*, **213**, 88-98.

Seither-Preisler, A., Patterson, R.D., Krumbholz, K., Seither, S. & Lütkenhöner, B. (2006b) From noise to pitch: Transient and sustained responses of the auditory evoked field. *Hearing Research*, **218**, 50-63.
Shahin, A.J., Roberts, L.E., Miller, L.M., McDonald, K.L. & Alain, C. (2007) Sensitivity of EEG and MEG to the N1 and P2 Auditory Evoked Responses Modulated by Spectral Complexity of Sounds. *Brain Topography*, **20**, 55-61.
Shapley, R. (1990) Visual sensitivity and parallel retinocortical channels. *Annual Review of Psychology*, **41**, 635-658.
Shapley, R. (2002) Neural mechanisms for color perception in the primary visual cortex. *Current Opinion in Neurobiology*, **12**, 426-432.
Singer, W. (1999) Neuronal synchrony: A versatile code for the definition of relations? *Neuron*, **24**, 49-65.
Singer, W. & Gray, C.M. (1995) Visual feature integration and the temporal correlation hypothesis. *Annual Review of Neuroscience*, **18**, 555-586.
Snyder, J.S., Alain, C. & Picton, T.W. (2006) Effects of attention on neuroelectric correlates of auditory stream segregation. *Journal of Cognitive Neuroscience*, **18**, 1-13.
Tallon-Baudry, C. & Bertrand, O. (1999) Oscillatory gamma activity in humans and its role in object representation. *Trends in Cognitive Sciences*, **3**, 151-162.
Tata, M.S., Prime, D.J., McDonald, J.J. & Ward, L.M. (2001) Transient spatial attention modulates distinct components of the auditory ERP. *NeuroReport*, **12**, 3679-3682.
Tata, M.S. & Ward, L.M. (2005a) Early phase of spatial mismatch negativity is localized to a posterior "where" auditory pathway. *Experimental Brain Research*, **167**, 481-486.
Tata, M.S. & Ward, L.M. (2005b) Spatial attention modulates activity in a posterior "where" auditory pathway. *Neuropsychologia*, **43**, 509-516.
Tian, B., Reser, D., Durham, A., Kustov, A. & Rauschecker, J.P. (2001) Functional specialization in rhesus monkey auditory cortex. *Science*, **292**, 290-293.
Tiitinen, H., Sinkkonen, J., Reinikainen, K., Alho, K., Lavikainen, K. & Näätänen, R. (1993) Selective attention enhances the auditory 40-Hz transient response in humans. *Nature*, **364**, 59-60.
Ungan, P., Şahinoğlu, B. & Utkuçal, R. (1989) Human laterality reversal auditory evoked potentials: stimulation by reversing the interaural delay of dichotically presented continuous click trains. *Electroencephalography and Clinical Neurophysiology*, **73**, 306-321.
Valtonen, J., May, P., Mäkinen, V. & Tiitinen, H. (2003) Visual short-term memory load affects sensory processing of irrelevant sounds in human auditory cortex. *Cognitive Brain Research*, **17**, 358-367.
von der Malsburg, C. & Schneider, W. (1986) A neural cocktail-party processor. *Biological Cybernetics*, **54**, 29-40.
Whittingstall, K. & Logothetis, N.K. (2009) Frequency-band coupling in surface EEG reflects spiking activity in monkey visual cortex. *Neuron*, **64**, 281-289.
Yost, W.A. (1996) Pitch of iterated rippled noise. *Journal of the Acoustical Society of America*, **100**, 511-518.

Yuval-Greenberg, S., Tomer, O., Keren, A.S., Nelken, I. & Deouell, L.Y. (2008) Transient induced gamma-band response in EEG as a manifestation of miniature saccades. *Neuron*, **58**, 429-441.

Zeki, S., Watson, J.D.G., Lueck, C.J., Friston, K.J., Kennard, C. & Frackowiak, R.S.J. (1991) A direct demonstration of functional specialization in human visual cortex. *Journal of Neuroscience*, **11**, 641-649.

Zipser, K., Lamme, V.A.F. & Schiller, P.H. (1996) Contextual modulation in primary visual cortex. *Journal of Neuroscience*, **16**, 7376-7389.